Charles University
Faculty of Science

Study programme: Biology
Branch of study: Biology

David Novák

Human lymphopoiesis and its examination via single-cell analysis

Lidská lymfopoéza a její výzkum pomocí single-cell analýzy

Bachelor's thesis

Supervisor: doc. MUDr. Tomáš Kalina, Ph.D.

Prague, 2018

Acknowledgement

I would like to thank Tomáš Kalina for being an excellent supervisor and for giving me an opportunity to educate myself in new ways at the CLIP laboratories in Prague. I also owe a debt of gratitude to Jan Stuchlý for introducing me to new data analysis approaches and providing highly useful feedback. Thanks to Jiří Gabriel for giving me and my friends the early chance to take a peek into the world of microbiology at the Institute of Microbiology and to the fantastic people from the laboratory of Martin Horák at the Institute of Physiology for allowing me to gain new pracical experience and tolerating my initial clumsiness. Moreover, I am grateful to Magda and to friends and family for valuable support, which came in many different forms.

Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 4. 5. 2018 David Novák

Abstract in English

Development of human B-lymphocytes is a convoluted process. A self-renewing stem cell progenitor in a primary lymphoid tissue commits to the lymphoid lineage. Subsequent B-lineage commitment entails somatic gene recombination processes which lead to the eventual expression of a surface antigen receptor. Functionality of the B-cell receptor, as well as successful testing for autoreactivity by the cell, are preconditions for the differentiation of a mature B-lymphocyte. Processes within this development are often investigated using single-cell analysis via flow cytometry, fluorescence-activated cell sorting and mass cytometry. Coupling these high-throughput methods with modern approaches to data analysis carries enormous potential in revealing rare cell populations and aberrant events in haematopoiesis.

Keywords: B-lymphocyte, lymphopoiesis, flow cytometry, FACS, mass cytometry, cluster analysis, FlowSOM, PCA, t-SNE, Wanderlust.

Abstrakt v češtině

Vývoj lidských B-lymfocytů je spletitým dějem. Samoobnovující se kmenový progenitor v primární lymfoidní tkáni je nejdříve předurčen k vývoji v lymfoidní linii. Vývoj B-buněčnou cestou pak zahrnuje somatické genové rekombinace, které vedou k expresi povrchového antigenního receptoru. Funkčnost B-buněčného receptoru a podrobení buňky testování na autoreaktivitu jsou podmínkami pro diferenciaci ve zralý B-lymfocyt. Děje v rámci tohoto vývoje jsou často zkoumány pomocí single-cell analýzy skrze průtokovou cytometrii, fluorescencí aktivované třídění buněk a hmotnostní cytometrii. Propojení těchto vysoce výkonných metod s moderními přístupy k analýze dat skýtá obrovský potenciál v nacházení vzácných buněčných populací a odchylných událostí v rámci krvetvorby.

Klíčová slova: B-lymfocyt, lymfopoéza, průtoková cytometrie, FACS, hmotnostní cytometrie, shluková analýza, FlowSOM, PCA, t-SNE, Wanderlust.

Contents

1 Introduction

The topics of this thesis are human lymphopoiesis and its examination using single-cell approaches. I focus on fate decisions and checkpoints during the development of B-lymphocytes from progenitor cells. Flow and mass cytometry are discussed as highly useful tools for gathering data which help immunologists understand the dynamics of how cells within the B-lineage evolve. Some modern approaches to the analysis of cytometric data are discussed and evaluated as an essential part of the overall narrative.

Flow cytometry is an exciting technique which allows for measurement of size, granularity and fluorescence from tags which are added to a cell. Both surface and intracellular markers can be used to identify different cell populations. Moreover, fluorescence-activated cell sorting can be used to physically sort cells according to the data measured.

Mass cytometry is a newer technique which addresses the issue of spectral overlap and disentangling emission spectra (compensation) when measuring fluorescence. Using metal-tagged antibodies, the panel of markers can be expanded radically.

However, processing datapoints which exist in many dimensions is not straightfoward. A number of fascinating algorithms for the automated visualisation and clustering of cell populations exists; some are presented in this work.

The development of B-lymphocytes (B-cells) in humans is among the processes which cytometry has helped to elucidate. Starting in fetal liver or adult bone marrow, stem cells differentiate into specialised cells, interacting with their surroundings, transcribing different genes. After commitment to the B-cell lineage, machineries for somatic gene recombination are assembled and immunoreceptors are gradually built. An immature and naïve B-cell with a functional receptor, after testing for autoreactivity, can then migrate to secondary lymphoid organs and play a role in humoral and cell-mediated immunity.

I created this work because I believe in the usefulness of gathering large amounts of data and using automatised computational analysis to uncover developmental trajectories of cells. It seems essential for life sciences to explore algorithmisation and machine learning to reveal these trajectories. For biologists to disregard this type of work would mean conceding one of the most exciting parts of scientific inquiry to others.

1.1 Aims

This work aims to evaluate the use of flow and mass cytometry for understanding human haematopoiesis, in particular human B-cell lymphopoiesis, and to describe B-cell lymphopoiesis. Moreover, it strives to present some principles of data analysis, as they pertain to processing cytometric data relevant to research into haematopoiesis.

2 Single-cell analysis

Analysis of single cells provides opportunities for examining differential expression of various genes and presence of markers. With increasing numbers of parameters which can be measured with single-cell technologies such as flow cytometry, fluorecence-activated cell sorting and mass cytometry, interpreting high-dimensional data is of great interest. These methods need to be addressed prior to discussing B-lymphopoiesis, as I will refer to them throughout.

2.1 Flow cytometry and cell sorting

Let us evaluate the principles which underly cytometry. A full historical overview is not in order, yet it is reasonable to review some older techniques also. Flow cytometry originated from a cell separator described by Fulwyler (1965). The apparatus would vibrate a cell suspension in a conducting medium to isolate single cells in droplets, so as to enable measurement of distinct events. The droplets would emerge in quick succession from a fluid jet. An electrostatic field between two deflection plates would then sort them according to volume.

The charging pulse needed for correct deflection would be determined beforehand, using the principle of a Coulter counter: before exiting through the jet, each cell in the medium would pass through a narrow orifice, causing a drop in current proportional to the volume of conducting fluid blocked from passing through. This drop would, in turn, be proportional to the volume of the cell (patented by Hogg and Coulter, 1971).

Flow cytometry added the ability to measure light scatter (which is pertinent to size and granularity) and fluorescence properties of single cells (or other particles) in the analyte (described in Van Dilla et al., 1969; reviewed by Adan et al., 2017). Two components are at play: the *fluidics* (inspired by Fulwyler) and the *optics*, coupled with detectors to acquire data of interest. If the machine retains the ability to sort cells, this is most commonly done based on fluorescence detection, termed *fluorescence-activated cell sorting* (FACS; Herzenberg et al., 1976). As opposed to the Fulwyler separator, the deflection parameter is a function of fluorescence, not volume.

2.1.1 Fluidics

A suspension of particles is injected into a stream of sheath fluid (most often phosphate-buffered saline). The coaxial flow of the analyte and sheath fluid (i.e. the flow along a common axis) aligns the particles in a single file. The stream of cells remains at the center of the flow cell. This focusing occurs due to the sample stream being injected at a greater pressure than that of the other concentric streams; the sample does not mix with

the surrounding sheath fluid (reviewed by Wilkerson, 2012).

2.1.2 Optics

The optical system includes a light source (often a laser) and other components for redirection and collection of light. The light is focused by a lens before reaching the analyte, and a set of optical filters and mirrors is used to deliver light of different wavelengths to photomultipliers and detectors.

Fluorescence and two light scatter properties are measured. The forward scatter parameter (FSC) measures the amount of light which travels forward, further along the path toward and past the analyte, allowing for only small change in angle. The side scatter parameter (SSC) measures the amount of light reflected upon reaching the cell. The former is a correlate of cell size, the latter of granularity in the cell structure (summmarised in Ormerod, 2008). A schematic representation of a flow cytometer can be seen below (figure 1).

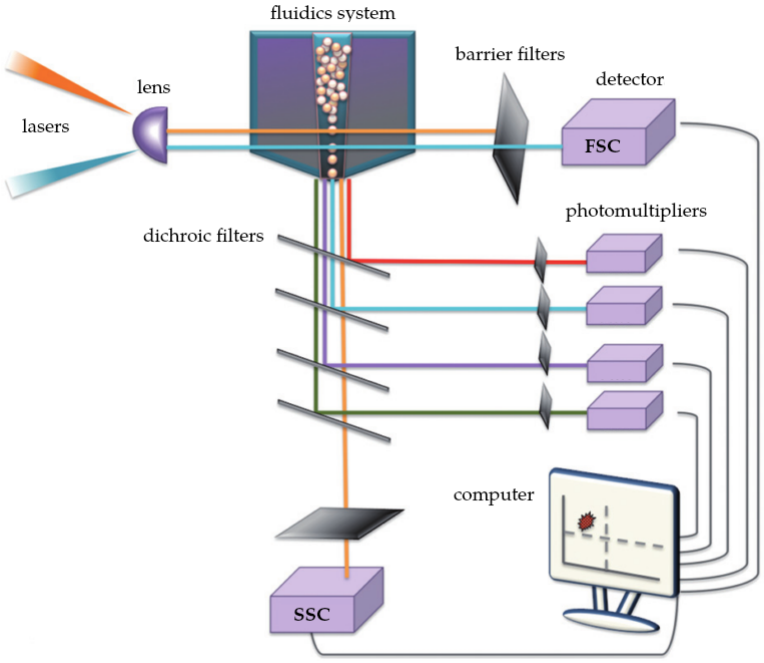

Figure 1. *Schematic representation of a generic flow cytometer, adapted from Adan et al. (2017).*

2.1.3 Fluorescent probes

The light source in a cytometer provides energy for electron excitation in the atoms of a fluorochrome. Subsequently, the electrons return to their ground state, losing energy

and emmitting a quantum of light. Thus, the optics of a flow cytometer measure both the autofluorescence and emission of light by fluorescent dyes added to detect surface or internal markers characteristic of particular cell types (Adan et al., 2017).

In an early paper which describes flow cytometry (as *cell microfluorometry*; Van Dilla et al., 1969), *Feulgen staining*, based on hydrolysis of DNA, was used to analyse the cell cycle of Chinese hamster ovaries by viewing a histogram of fluorescence intensity at different wavelengths. Here, peaks in signal were found to correlate with the G1 (diploid) and G2 (tetraploid) phases of the cycle.

Later, nucleic-acid intercalators like ethidium-bromide (also used in gel electrophoresis) were used to monitor DNA content to understand kinetics of adenoviral infection (Lemay and Collyn-d'Hooghe, 1984). DNA replication caused by the infection could be detected as an increase in signal and was compared with a mock infection by a DNA-negative control virus.

Morevover, the intercalating 4',6-diamidino-2-phenylindole (DAPI), binding to minor grooves in DNA, is often used. First employed to isolate mitochondrial DNA (Williamson and Fennell, 1974), DAPI is currently used in some panels as a marker for dead cells, as it cannot permeate the membrane of a live cell. For instance, Sonn et al. (2016) include DAPI for identifying circulating tumor cells, demonstrating usefulness of DAPI in an oncological study.

Simmons et al. (1991) used various mouse anti-human antibodies (primary Ab) and goat anti-mouse immunoglobulin G (IgG; secondary Ab) conjugated with fluorescein isothiocyanate (FITC; one of the commonly used dyes) which would bind to CD34–a marker of lymphocyte precursors in the bone marrow (see section 3). This is an example of primary and secondary antibody use in immunochemical staining.

Dead cells can be detected by incorporating another fluorescent intercalator, propidium iodide (PI), into DNA. This was used by Kikushige, et al. (2008). In addition, a range of markers to detect progenitors in the myeloid and lymphoid lineage of cells (important in the immune system; see section 3) was added. For instance, an anti-hFlt-3 antibody was used to mark a receptor tyrosine kinase associated with myeloid progenitors, which go on to become erythrocytes, granulocytes, megakaryocytes or monocytes, partly due to the signalling enabled by this receptor. The antibody was stained with FITC, emitting green light. This is an example where the fluorochrome is bound directly to the primary antibody, with no need for a secondary antibody.

Some laboratories use immunoglobulins conjugated with the water-soluble biotin (vitamin B_7). At basic pH, biotin substitutes an amino-group of one the amino-acids which make up the Ig. The biotinylated antibody acquires high affinity for avidin, a tetrameric protein which forms stable covalent bonds with biotin. This system is often employed in immunohistochemistry (Bratthauer, 2010).

Biotinylation was used by Garvy and Riley (1994), who focused on the role of the cytokine interleukin-7 (IL-7) in B-cell development. In order to stain for IL-7 receptors (see section 3.2), the team used biotinylated IL-7 conjugated with avidin-linked phycoerythrin (PE). Furthermore, in a nine-parameter immunophenotyping study by Bigos et al. (1999), the authors boosted efficiency of their CD1 staining (marking MHC class I molecules) by using a biotin-streptavidin link. Streptavidin, boasting a lower propensity for unwanted non-specific binding, was conjugated with a *Texas Red* dye. The use of biotin and avidin or streptavidin showcases the opportunities to strengthen the staining of cell markers.

2.2 Mass cytometry

Mass cytometry was first described by Bandura et al. (2009) as a novel technique for multitarget immunoassays. Based on mass spectrometry, it would reap the benefits of using plasma for high throughput and prove itself useful for various analyses.

In contrast with classical flow cytometry, mass cytometry uses elemental metal tags to detect distinct cellular markers, providing a narrower bandwidth of emission. This is to reduce spectral overlaps and do away with the need for data compensation (see section 2.3.1).

As reported in cytometric studies where many parameters of interest are examined simultaneously, the issue of spectral overlaps is the principal factor limiting maximum numbers of parameters and interpretability of data (Roederer, 2001).

To understand mass cytometry, we need to review the technologies which precede it. One of the approaches to secure a stable and fast throughput of cells is to vaporise the analyte and inject it into a stream of plasma. (Plasma is a state of matter where the behaviour of gaseous ions is mainly governed by long-range electric and magnetic fields.) For *inductively coupled plasma* (ICP), the energy for the gradual phase transition from gas to plasma is supplied by electromagnetic induction. Nomizu et al. (2002) applied these principles to detect airborne zinc particles by recording a signal via electron multipliers. Prior to this, the group (1994) demonstrated that individual cells can also be injected into this stream to detect endogenous calcium.

Becker et al. (2007) describe the use of a nebuliser to create aerosol from biological tissue and using argon as a carrier gas. The analyte was delivered to an ICP source and single ions were detected and sorted by varying mass-to-charge (m/z) ratios. The m/z variable, an important value in mass spectrometry, is the focus of interest for the researcher, who seeks to identify abundances of ions of particular mass-to-charge ratio in a given sample.

One of the approaches to obtaining a mass spectrum in a spectrometer is to accelerate the ionised particles and deflect them in a magnetic field. With changes in energy that

is supplied to an electromagnet, the severity of deflection is modified and disparate sets of ions are delivered to a detector (described in Glish and Vachet, 2003). With time-of-flight (TOF, time until detection of particle) and the m/z parameter correlated, we can employ these principles in a generic *CyTOF* (Cytometry-by-Time-of-Flight) mass cytometer, measuring time until detection of different metal tags.

2.3 Data processing

Processing data gathered by the aforementioned techniques is an interesting topic to explore. Computer analysis allows the researcher to manipulate data manually, as well as to employ automation. The latter option is of growing importance.

In order to increase reproducibility of experimental results, the common FCS (*Flow Cytometry Standard*) file format for storing cytometric data was proposed in 1984 by Murphy and Chused. The FCS format records fluorescence and light scatter values, as well as time parameters and other information for the purpose of annotation and preserving changes made to the raw data by processing. The FCS 3.0 standard allowed for larger datasets and extended support for cluster analyses (Seamer et al., 1997).

2.3.1 Compensation

Researchers are often confronted with overlapping emmission spectra in their experimental panel of fluorochromes. This causes a signal from one fluorochrome to appear in multiple channels. Compensation seeks to disentangle these spectra. Parks et al. (1977) were the first to achieve this, with a protocol to compensate overlaps from fluorescein and rhodamine signals. In the experiment, expression of two different antigens on IgM and IgD molecules (different isotypes of antibody) from murine spleen was investigated. Prior to this experiment, the two markers on the mouse lymphocyte sample would have had to undergo separate analyses. (An example of this, rather elaborate, approach was presented by Curbelo et al. /1976/. In their design, cells would flow through a capillary and up to 8 distinct optical measurements would be acquired. /A system resemblant of *microfluidics*– another prominent single-cell technology./)

The Parks paper presents a novel approach which permitted the team to measure their two signals at a single interrogation point for each cell and to unravel the data during measurement. A fluorochrome-specific and filter-specific constant needed to be determined empirically to quantify the size of interference between the channels. Measurements were corrected by added electronics which would divert some signal to the other detector, taking into account the differences in amplification between channels.

Nonetheless, the system proved inefficient, as the circuitry needed to perform real-time compensation grew quickly in size with each new parameter and the setting yielded

poor results with more than 4 parameters. Bagwell and Adams (1993) tackled the issue of scalability with a software approach, where a matrix showing crossovers between any two parameters is calculated.

Roederer et al. (2001) conducted compensations on a CD8-Cy5-Phycoerythrin conjugate (staining mainly *Tc* cells, emmitting far-red light) and CD4-Phycoerythrin (staining mainly *Th* cells, emmitting yellow light). The team showed further advantages of software-based data manipulation. Ceasing to measure fluorescence on a logarithmic scale (which is used for visualisation of fluorescence parameters), they would measure data on a linear scale and transform it later. This proved to minimise errors.

2.3.2 Gating

Gating, the principal step in cytometry data analysis, is the act of separating a population of cells based on light scatter properties or differential expression of markers. An example of a gating strategy applied to a dataset can be seen below (figure 2).

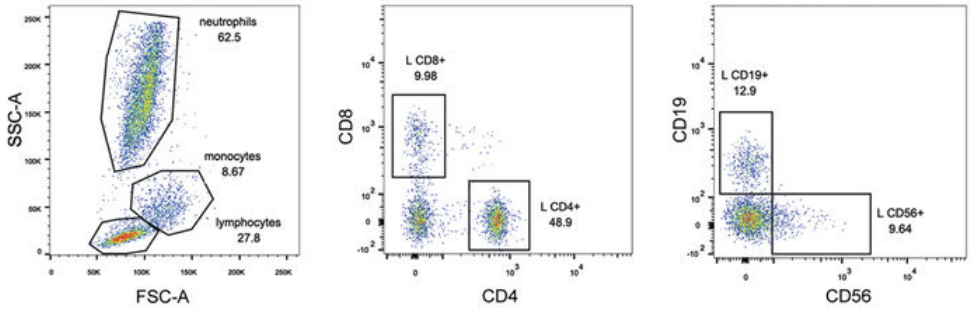

Figure 2. *A partial gating strategy adapted from Lechner et al. (2015), illustrated using dot-plots. FSC-A and SSC-A refers to forward-scatter and side-scatter area, i.e. the integral of signal intensity as a function of time. Light scatter parameters are displayed on a linear scale, others are log-scaled.*

2.3.3 Dimensionality reduction via PCA and *t*-SNE

Analysing cytometric data using 2-dimensional dot-plots is constraining. For instance, to reveal the relationships between *each* pair of parameters in a 12-parameter panel, 66 two-dimensional dot-plots must be constructed (i.e. the sum of all positive integers smaller than 12). By this logic, a 30-parameter mass cytometric panel would warrant 435 dot-plots.

To tackle the issue of multiple dimensions, some techniques allow projections of data such that dimensionality is reduced to 2 or 3 dimensions. *Principal component*

analysis (PCA) considers a matrix X of fluorescence intensities per event. Very generally, this is reinterpreted as a product of *(1)* some matrix of baseline values and *(2)* another matrix explaining variation versus the baseline. The second matrix is built by identifying which non-correlated linear combinations of the original variables (rows/columns of X) contribute to the variance within this dataset the most. Only these newly calculated parameters (called *principle components*) are then used in visualising the data (Wold et al., 1987).

Geometrically, principle components can be imagined as multiple lines of best fit through the dataset. If we imagine our dataset as a 3-dimensional 'cloud' of points, the first PC intersects the cloud along its longest axis, the second PC is perpendicular and the third PC is orthogonal. The two or three axes of a resulting graph will correspond to these components and all the data will be mapped into this space. The same principle of constructing PCs as lines through a cloud of data can be applied to datasets of higher dimensionality. Some residual variation, not explained by the approximation provided by PCA, will be present in the original matrix – in an ideal situation, this will only be noise in the data.

Lugli et al. (2007) designed an 8-colour panel to investigate differences in the T-lymphocyte compartment between donor age groups. The team gated data based on light scatter properties (lymphocyte gate), CD3 expression (marking T-lymphocytes) and CD4 versus CD8 (*Th* vs *Tc* cells). Thereafter, other markers, including CD127 (an interleukin-7 receptor subunit, involved in survival), CD95 (the first apoptosis signal receptor, involved in cell death) or CD38 (a glycoprotein involved in cell adhesion and activation), were used to create 'positive' and 'negative' gates (or, with CD38, 'negative', 'bright' and 'dim'). Groups of cells based on *(1)* donor group and *(2)* combination of positivities/negativities were then subjected to PCA, yielding three principal components (PCs). The study elaborates on the use of various data normalisation methods used prior to the analysis. Figure 3 shows one of the PCA-generated graphs from the study.

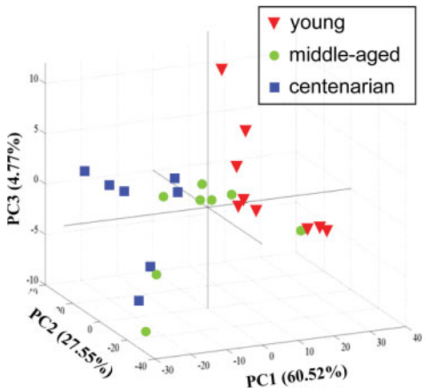

Figure 3. *An example of a 3-d plot based on PCA on $CD4^+$ T-cells (adapted from Lugli et al., 2007). Distribution of cell populations on a principal component (PC) scales are*

displayed, with metrics of influence attributed to each component. Each PC is a linear function of the measured fluorescence parameters.

After observing that groups of cells were mapped to different positions based on donor age, the team would report on the most striking differences (for instance, with regard to $CD4^+$ cells, a $CD127^+CD95^-CD38^{dim}$ phenotype was identified as largely characterising young donors and a $CD127^-CD95^+$ phenotype seemed typical for cenentarians – people over 99 years of age).

A more prevalent technique for dimensionality reduction, currently of great importance to cytometrists, is t-distributed stochastic neighbour embedding (t-SNE). Van der Maaten and Hinton (2008) address the issue of assuming linearity, warning that techniques including PCA 'focus on keeping the low-dimensional representation of *dissimilar* datapoints *far apart*', yet it is often 'more important to keep the low-dimensional representations of *very similar* datapoints *close together*.' Moreover, they introduced t-SNE: an algorithm for visualising data while preserving much of the local structure and some of the global structure of high-dimensional data.

The original stochastic neighbour embedding algorithm, introduced by Hinton and Roweis (2002), uses a probabilistic approach to determining neighbourhood of events. This means that for each point i in the dataset, a 'probability' is computed that it picks any other one point j as its neighbour. This is achieved by imagining a straight line connecting the two points. Then, a Gaussian (bell-shaped) curve is drawn along the line, so that the center of the curve (the peak) is at i. Dissimilarity between the points increases with decrease in the height of the curve at j (i.e. value of the Gaussian function at j). The width of the Gaussian curve is manipulated with each new i according to local entropy: generally, the denser the region is, the narrower the curve is, and vice versa. To map all the datapoints in a space with fewer dimensions, they are mapped stochastically, then rearranged. Based on dissimilarities in the original space, dissimilar points repel, whereas similar points attract.

The t-SNE algorithm (Hinton and Roweis, 2002) makes some changes to the original SNE algorithm, including the use of t-distribution instead of Gaussian distribution. Figure 4 shows a typical t-SNE graph. The problem of 'overcrowding' is inherent for t-SNE, with the reduction of dimensions resulting in space constraints. A novel method of *hierarchical stochastic neighbour embedding* (HSNE) offers the capability to run multiple SNE analyses in iterations, identifying representative 'landmark' cells and passing them as input to another SNE level. This reduces the need for downsampling, and the analysis boasts good performance in identifying rare immune cell types, as tested on samples from patients with gastrointestinal disorders (Van Unen et al., 2017).

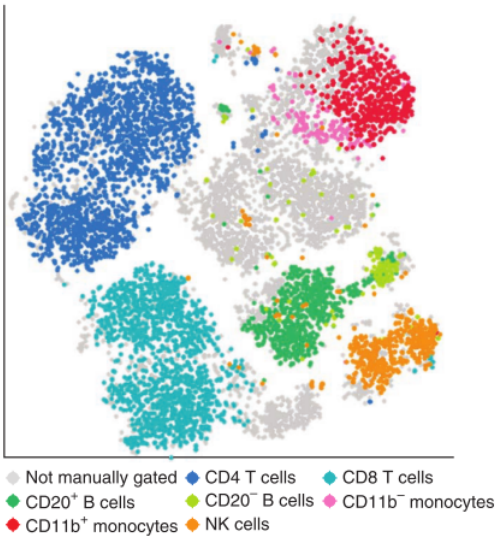

Figure 4. *An sample output of viSNE (software implementation of t-SNE), as applied to a mass cytometric analysis of healthy human bone marrow, with stains for 13 markers. Some populations gated manually by an expert (in bright colours) were clearly separated into distinct regions by the t-SNE algorithm itself. Adapted from Amir et al. (2013).*

2.3.4 Cluster analysis using self-organising maps

Many different algorithms are available for automatised identification of clusters of events within a dataset (some are reviewed by Weber and Robinson, 2016). These can be used as tools for identifying cell populations. For the sake of brevity, a single implementation, that of *self-organising maps* (SOMs), is discussed here.

Van Gassen et al. (2015) set out to implement a fast technique for both visualisation and interpretation of data from cytometry. Interestingly enough, they cite a study (Kohonen, 1998) which describes an architecture for mapping large amounts of textual documents in English as points on a grid. Kohonen communicates a 'self-teaching' algorithm which creates a *self-organising map*, accounting for similarities between representative samples of text from each document. Each sample is a triplet of words, where a word is represented by a 90-dimensional vector; a triplet is then termed a *'contextual pattern vector'* (an equivalent to this would be to a hypothetical 270-parameter panel of measurements in our context).

The algorithm which clusters these 'fingerprints' of documents is explained by the Van Gassen team. A grid of nodes (equivalent to random points in the dataset) is created, where each node corresponds to some point in a space (where dimensionality equals the number of measured parameters). Individual points from the dataset (contextual vectors or individual cell measurements) are then assigned to the closest neighbouring node. This node and all its neighbour nodes which find themselves closer than a set value (ε) shift

toward the newly processed point by a distance determined by a *learning factor* (α). As the algorithm is fed more points and the map is trained, both ε and α decrease.

As Kohonen elaborates, the finished map can be used to separate the given space into regions by deeming the nodes *seeds* to which we assign regions (*bins*) consisting of events closest to the particular node.

In contrast, the Van Gassen team presents *FlowSOM*, a tool for cytometry, which constructs a *minimal spanning tree* as output of the analysis (shown in figure 5). In this structure, the nodes are interconnected in such a way that the weights of branches (connections) are minimised. The team advises users of FlowSOM to select a number of clusters that is higher than the expected number of gates one would acquire by manual gating. This way, previously unknown cell populations or transitioning cells can be captured in the output. Ultimately, a *meta-clustering* process can be carried out, whereby the most similar of the many clusters produced are merged.

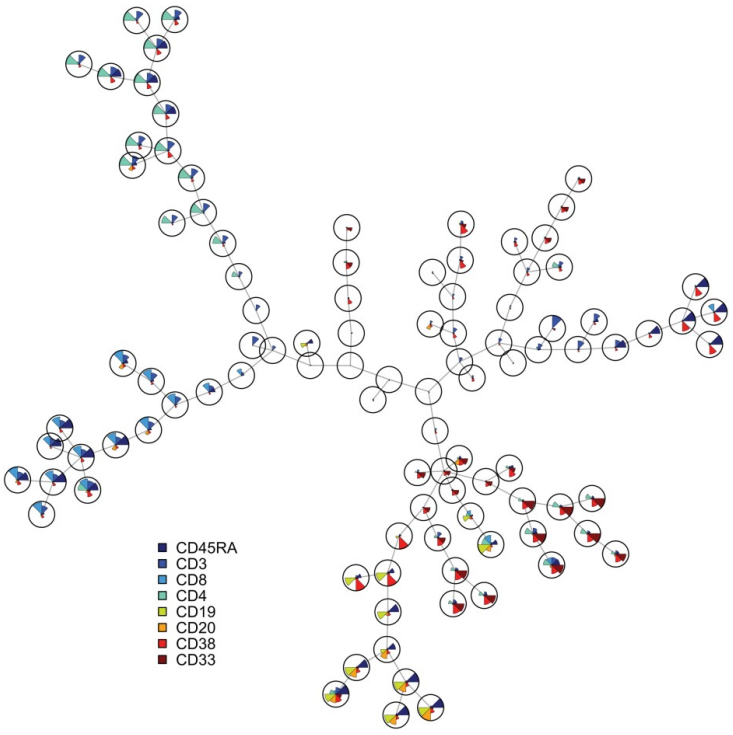

Figure 5. *A minimal spanning tree as output of FlowSOM applied to a bone marrow dataset (Van Gassen et al., 2015). With a pie chart displayed for every node, the displayed relative representation of various markers can serve to identify various cell populations.*

The authors of FlowSOM applied the algorithm to a human bone marrow (BM) mass cytometry experiment. Staining for 31 markers, two of the samples used were one of unstimulated BM cells and one of cells stimulated by IL-3 (a factor in maturation of progenitors). After a manual gating of single cells, 13 surface markers were used to build the map. Various cell populations (including naïve and memory T-cells and B-cells, NK

cells, myeloid populations, different progenitor stages, etc.) were identified. FlowSOM also allowed the team to look at elevation in some intracellular markers (the pSTAT5 and pSTAT3 signal transducers) that was caused by IL-3 stimulation in the compartment capturing most *plasmacytoid dendritic cells* (pDCs; found in peripheral lymphoid organs). This may serve as some evidence that the clusters created by FlowSOM were sensible, as the algorithm was not fed data pertaining to these intracellular markers, yet it uncovered a population which would elicit a response via these markers.

Due to its relatively short runtimes and the ability to detect obscure cell populations, I find FlowSOM particularly interesting as a method to extract information from large mass cytometry datasets, related, for instance, to haematopoiesis and its disorders.

3 B-cell lymphopoiesis

Lymphopoiesis is the differentiation of immunocytes of the lymphoid lineage. These include B- and T-lymphocytes and natural killer (NK) cells, all important for the ability of the human immune system to function (explained in Horejsi et al., 2017).

B-cells (B-lymphocytes) are a subset of immunologically relevant cells with roles in both humoral and cell-mediated immunity. They express surface immunoglobulin (antibodies) with a wide range of specificities which allow them to recognise foreign antigens in the body. B-cell development takes place in the bone marrow, with mature B-cells migrating to secondary lymphoid organs (eg. spleen, lymphatic nodes) where they are stimulated by foreign antigens, diversify their repertoir of antibodies and differentiate into different subsets, including memory B-cells and plasma cells (this is reviewed by LeBien and Tedder, 2008).

The result of successful B-lymphopoiesis is establishment of a repertoire of naïve cells which generally do not respond to stimulation to self-antigens in the body, but are ready to react to contact with a foreign antigen. For this to happen, progenitors must be stimulated by adhesive molecules and cytokines, invoking signalling pathways and transcriptional factors which, in turn, cause the expression of different genes. This includes the expression of molecular machinery which rearranges immunoglobulin gene segments, leading to the creation of B-cell receptors that can have different specificities. This occurs through an interesting process of introducing double-strand breaks to DNA, which is usually a sign of damage, and joining non-homologous DNA coding ends.

The following chapters describe the differentiation of a haematopoietic stem cell (HSC) to a common lymphoid progenitor (CLP) and its development into a pro-B cell (which initiates somatic recombination of immunoglobulin heavy chain genes) and pre-B cell (which expresses a pre-B cell receptor with a surrogate immunoglobulin light chain). Moreover, the creation of an immature B-cell with a functional B-cell receptor and its testing for autoreactivity is addressed. It is shown that cytometry and cell sorting play a big role in reasearch into these processes. Ultimately, the role of mass cytometry in uncovering the developmental trajectories of human B-lymphocytes is evaluated.

3.1 Origin of the common lymphoid progenitor

Lymphopoiesis begins with the differentiation of haemotopoietic stem cells (HSCs) in fetal liver or adult bone marrow into cells which are committed to lymphoid lineage (*common lymphoid progenitors*).

For identification of lymphoid progenitor cells which emerge from HSCs, the CD34 surface protein marker is used extensively. First described through the binding of a murine

antibody and subsequent cell sorting (Katz et al., 1985), CD34 had later been detected to also mark stromal cell populations of the marrow (Simmons and Torok-Storb, 1991). It was later found that in murine lymphopoiesis, CD34 mostly marks short-term non-renewing multipotent progenitors downstream from HSCs; a $CD34^{-/low}$ phenotype (Osawa et al., 1996) with positivity in CD38 (a glycoprotein with adhesion and signalling functionalities) would then act as a better indicator for purifying HSCs in FACS analyses (Morrison and Weissman, 1994). Yet in humans, CD34 expression was succesfully used for gating long-term renewing HSC populations (Okuno et al., 2002). Another marker of progenitor cells is the c-Kit (CD117) receptor tyrosine kinase: a proto-oncogene which mediates survival and proliferation by inactivating pro-apoptotic proteins in the cell (Edling and Hallberg, 2007).

Stromal cells in the bone marrow stimulate HSCs via adhesion and release of cytokines. Adhesion, chiefly via the VCAM-1 (vascular cell adhesion molecule-1) sialoglycoprotein, has been shown in mice through labelling by antibodies with radioactive iodine isotopes (Jacobsen et al., 1996) and the disruption of stromal cell-HSC aggregates by injection of anti-VCAM-1 antibodies (Funk et al., 1994).

Furthermore, adhesion via the transmembrane N-cadherin protein, as well as a stronger binding via CXCL12 (Burk et al., 2015) was demonstrated for humans on a mesenchymal stromal cell model. The presence of CXCR4, a corresponding chemokine receptor, has been verified via flow cytometry of $CD34^{+}$ progenitor cells (Möhle et al., 2013), where some $CD34^{+}$ leukaemic cells were also found unable to undergo migration across bone marrow endothelia.

Initially, HSCs differentiate into *multipotent progenitor cells* (MPPs), with continuous CD38 upregulation observed in developmental trajectories inferred from mass cytometry panels. (Bendall et al., 2014; Bagwell et al., 2015). As opposed to HSCs, the MPPs represent an oligopotent population with reduced capability of self-renewal; the creation of MPPs is associated with expression of the tyrosine kinase 3 (Flt-3; CD135) cytokine receptor in mice (Adolfsson et al., 2005). In humans, however, Flt-3 expression already occurs in the previous developmental stage—the HSC—as confirmed by flow cytometric analyses (Kikushige et al., 2008), where significant Flt-3 expression was observed both in earlier CD38-negative and later positive populations (figure 6 shows a partial gating strategy from this study).

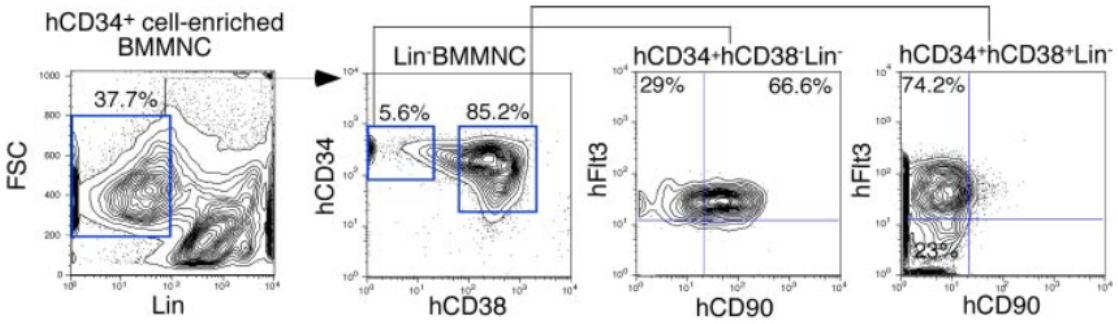

Figure 6. *A contour plot from Kikushige et al. (2008) showing a short gating strategy for $CD34^+$-sorted bone marrow mononuclear cell. Lineage-specific marker-negative cells are gated and CD38-negative and positive gates are applied. In addition to selection of significant human (h) Flt-3 populations, some negative correlation between CD38 and CD90 (a surrogate HSC marker, used as indicator of pluripotency) is apparent on the level of the given population, along with a greater portion of $hFlt\text{-}3^-$ events in the $CD38^+$ compartment.*

Subsequent differentiation into common lymphoid progenitor cells (CLPs) is intertwined with effects of interleukin-7 (IL-7), secreted by stromal cells. Stimulation of the IL-7R receptor is a feature which accompanies creation of CLPs, as opposed to myeloid progenitors (Johnson et al., 2005), where dose-dependent proliferation of purified $CD19^+$/IL-$7R^+$ cells occured upon treatment with IL-7, via induced phosphorylation of STAT5, involved in further signalling. Some researchers recognise *lymphoid-primed multipotent progenitors* (LMPPs) as a preceding stage in lymphoid development (Dias et al., 2008). LMPPs are characterised as cells which may give rise to lymphoid lineages and some myeloid lineages (granulocytes or macrophages, yet not erythroid cell subsets). These cells express Flt-3 and this expression may overlap with IL-7R expression, otherwise common for CLPs (Adolfsson et al., 2005). Expression of the E2A transcription factors is crucial for lymphoid specification (discussed later; Dias et al., 2009).

The expression of Flt-3 in LMPPs and later stages is at least partly enabled by the Ikaros transcription factor. Ikaros is a chromatin regulator with a protein-binding and DNA-binding zinc finger cluster (for dimerisation and activating transcription, respectively) in most of its isoforms (Molnár and Georgopoulos, 1994). Sun et al. (1996) describe multiple forms, created by alternate splicing of the Ikaros transcript–some reinforcing interaction with DNA and some repressing it by inhibiting other Ikaros (Ik) proteins.

In attempts to specify the point where T-lymphopoiesis and B-lymphopoiesis diverge, the Ly6D surface marker was identified as specifying a B-cell fate by a flow cytometric analysis (Inlay et al., 2009). By mapping this onto the HSC-to-MPP-to-LMPP-to-CLP timeline, most prospective T-cells were observed to branch out during the LMPP phase of development. Among all CLPs, the Ly6D marker would enable identification of those progenitors which would produce B-cells exclusively, as opposed to those which were un-

restricted as to which lymphoid population they would create.

3.2 Commitment to the B-lineage

The common lymphoid progenitor later commits to development into B-cells. To unravel the main processes that lead up to this commitment, let us first evaluate the significance of an important cytokine, interleukin-7, and its signalling at this stage.

The impact of interleukin-7 on human B-cell development is controversial and diverse. In a 1996 work by Pribyl et al., a human bone marrow $CD34^{+}/CD19^{-}$ cell culture (where CD19, a prominent signalling adaptor, marks most B-cell stages) was isolated using FACS to demonstrate the ability to commit to the B-lineage without IL-7. (Crucially, this cell culture originated from a fetus.) This is in contrast with the findings of Peschon et al. (1994) who used flow cytometry to demonstrate deficiency in cells expressing heat stable antigen (HSA/CD24; B-cell marker) in adult mutant mice.

Fetal B-cell development is distinct in that other cytokines, including thymic stromal lymphopoietin (TSLP), produced by various stromal cells and epithelia, may seem to fulfill some approximate roles of IL-7 (Voßhenrich et al., 2003). TSLP promotes development of immature B-cells by both fetal liver and adult bone marrow cell cultures in mice. Furthermore, its effect is impeded by monoclonal antibodies against the IL-7Rα chain of the IL-7 receptor dimer, whereas the same approach to block the other (γc) chain of the receptor shows no effect (Levin et al., 1999). The same mechanism seems to be responsible for the apparently normal B-cell counts in patients with *X-linked severe common immunodeficiencies* (XSCIDs) lacking a functional γc chain of IL-7R (Puel et al., 1998), where TSLP was suggested as a substitute for IL-7 in adult progenitors (as tested on mice by Chappaz et al., 2007). The role of TSLP as this substitute, however, has been challenged since. Jensen et al. (2008) conducted several comparative analyses and identified the Flt-3 ligand as a more proximate actor in fetal B-lymphopoiesis. Moreover, findings on the role of TSLP in human B-cell development are scarce, with a singular *in vitro* study suggesting that TSLP contributes to the commitment of fetal progenitors, with the two chains of the TSLP receptor (IL7-Rα and TSLPR) expressed more in the bone marrow than the liver (Scheeren et al., 2010).

Luckily, a study by Parrish et al. (2009) managed to clarify the matter as it relates to IL-7 and FLT-3 in humans. Using human-only co-cultures of neonatal cord blood (CB) or adult bone marrow (BM) progenitors and human stroma, the work suggests more than 60-fold increases in $CD34^{+}$ and $CD34^{-}$ B-cell precursor counts in both CB and BM upon stimulation by IL-7. However, a consistent 'baseline' of $CD19^{lo}$ progenitor production was *only* observed in CB, while the *in vitro* generative capabilities of BM cells decreased along different stages of ontogeny, rendering IL-7 progressively more important in regular B-lymphopoiesis.

A number of signalling pathways (JAK-STAT, Src, PI3K/Akt, MAPK/Erk) has been observed in action as a result of IL-7 signalling, using some mouse and human models (as reviewed by Corfe et al., 2012). A plethora of transcription factors are implicated in futher stages of B-cell development.

One of the prominent transcription factors, implicated in assisting the transition to the CLP and the pro-B cell, is the PU.1 protein. PU.1-deficient mice (with a null mutation) die in early development; Schweitzer and DeKoter (2004) used flow cytometry to quantify the related changes in $CD19^{+}$ cell counts which document B-cell count reductions. While the mutants were deficient in pro-B cells, those transducted with an IL-7Rα retroviral vector restored some IL-7 signalling. Interestingly, Inomata et al. (2006) observed a negative correlation between expression of PU.1 and Flt-3 in mice with *acute myeloblastic leukaemia* (AML) where Flt-3 was over-expressed, showing the possible role of PU.1 as a repressor of an *flt3* promoter in the cell. In general, lower levels of PU.1 are observed to correlate with commitment to the B-lineage, while elevated levels are typical of myeloid lineage commitment, back at the LMPP stage of development (DeKoter et al., 2002).

The early B-cell factor (EBF) is a protein which promotes commitment to the B-lineage in particular, in synergy with E2A proteins. EBF acts both as an antagonist to pro-myeloid factors (including PU.1) and promoter of *Pax5*, an actor in repressing genes specific for other lineages (Pongubala et al., 2008), yet EBF can act independently of it.

The E12 and E47 transcription factors, created via alternative splicing from the E2A gene, also induce Pax5 transcription, and have been implicated in enabling transcription of the recombinase needed for later Ig segment rearrangement, the CD19 signalling adaptor and other important proteins needed for B-cell maturation (Bain et al., 1994), as inferred from observing deficient mice.

Moreover, in human T-cell lymphoma patients suffering from deletion of E2A, the *c*-Myc protoncogene, which usually remains down-regulated, and Cyclin-dependent kinase 6 (Cdk6), which regulates transition to S-phase in the cell cycle and the tumor suppressor retinoblastoma protein (pRB), are expressed at abnormally high levels (Steininger et al., 2011); this documents the role of E2A as a suppressor of tumor growth. Aberrant distributions of the proliferating cells by phases in the cell cycle were observed via flow cytometry: bromodeoxyuridine (BrdU), analogue of thymine, conjugated with allophycocyanin (APC), was incorporated into the DNA of malignant cells and staining with 7-Aminoactinomycin D (7-AAD; a DNA intercalator used for determining cell cycle phases) followed. One of the cell cultures was transfected with E47. Gating of cells by stages of the cell cycle was possible by distinuishing the various clusters through correlating DNA content with relative amounts of incorporated BrdU (indicative of DNA synthesis). An E47-stimulated culture showed a significantly smaller portion of cells in S-phase and a higher portion in the G0/G1 phases, corroborating the evidence from quantitative PCR pointing to E2A-DNA binding as a mechanism of repressing cancerogenesis.

3.3 The pro-B cell and its fate

The pro-B cell is the stage at which somatic recombination, leading up to the eventual expression of a pre-B cell receptor (a predecessor to the full B-cell receptor), is initiated.

The transition to pro-B cells can be identified by some increase in expression of CD19 and CD10 (and endopeptidase often used in diagnostics of cancerogenesis) and CD22 (a transmembrane lectin structurally related to immunoglobulin molecules, tracking development along the one-dimensional B-lymphopoietic continuum). A $CD34^+CD10^+CD19^+TdT^+$ phenotype (where TdT is the terminal transferase in the nucleus, see below) is characteristic for pro-B cells (Bagwell et al., 2015; Bendall et al., 2014).

Earlier, we referred to the Pax5 transcription factor as appearing 'downstream' of EBF and E2A. Pax5-deficient pro-B cells retain their potential to differentiate into myeloid cells and fail to develop into B-cells (Heavey et al., 2003; Urbánek et al., 1994). Holmes et al. (1992) report that Pax5-deficient cells fail to repress Flt-3 transcription, whereas after reintroducing Pax5, the pro-B cells retain this ability. This may elucidate the mechanism by which PU.1 deficiencies cause Flt-3 overexpression. Some pertinent data from immunodeficient patients are available with regard to Pax5 deficiency (Kaneko et al., 1998). As observed in this study, CD19-expression was found to correlate with Pax5, which corroborates the findings of Kozmik et al. (1992) who isolated Pax5 as a 'B-cell specific activation protein'.

The transition from a lymphoid progenitor to the pro-B cell stage is accompanied by the initial stages of DNA recombination, leading toward the eventual creation of a pre-B-cell receptor (pre-BCR). The expression of a recombinase is essential for the rearrangement of gene segments that precedes the formation of a pre-BCR. The recombinase activity gene-1 (RAG-1) was identified as a principal actor. The expression of RAG-1 along with RAG-2 is common to both B- and T-lineage progenitors (Schatz et al., 1989). Oettinger et al. (1990) found that transfects with RAG-2 showed increased frequencies of recombination.

The previously discussed Ikaros transcription factor was identified to induce expression of the RAG-1–RAG-2 complex [the V(D)J recombinase] in the following experiment. Reynaud et al. (2009) transfected some MPPs deficient in Ikaros with a retroviral vector carrying EBF. The EBF-transfected cells achieved, albeit delayed and limited, differentiation into $CD19^+$ precursors, whereas the group transfected with a control retrovirus went on to produce myeloid cells. (This rescue of B-lineage potential could not be achieved by E2A or Pax5, however.) Nonetheless, the generated $CD19^+$ cells were unable to express the *Rag1* and *Rag2* genes, as well as the *Dntt* gene, coding terminal deoxyribonucleotidyltransferase (TdT; discussed later).

Matthews et al. (2007) report on a particular conserved tryoptophan residue of RAG-2 which allows for the recognition of a nucleosome where histone 3 is trimethylated

at position-4 lysine (H3K4me3). The structure responsible for the RAG-2–H3K4me3 binding is a PHD finger (a cysteine and histidine residue-containing motif often involved in interactions with histones), as revealed in sequencing by Callebaut and Mornon (1998).

Prior to the rearrangement of gene segments in the *Igh* (immunoglobulin heavy chain) loci, these segments migrate from the transcriptionally repressed nuclear periphery to central regions of the nucleus in the early pro-B cell (Kosak et al., 2002), wherein IL-7 stimulation appears to amplify this process. The IgH gene consists of the V_H ('*variable*'), D_H ('*diversity*') and J_H ('*joining*' or '*junctional*') regions. A mechanism of looping within the gene for IgH brings the V region closer to the D/J cluster within the *Igh* structure, as observed via fluorescent in-situ hybridisation of the three regions by Sayegh et al. (2005). Hewitt et al. (2009) observed that homologues within the IgH loci across chromosomes pair up under the influence of RAG-1–RAG-2.

Subsequently, the V(D)J recombinase facilitates the joining of some D_H and J_H segments in the pro-B cell (preceding V_H-to-D_H joining). Critically, a few hundred V-segments, about 50 D-segments and 9 J-segments are present, ensuring an element of randomness in which of the segments are connected (Horejsi et al., 2017). This randomness, however, is kept at bay by a principle dubbed the *12/23 spacer rule*. As explained by Steen et al. (1996), the conserved *recombination signal sequences* (RSSs) contain a heptamer sequence adjacent to the coding segment and a 12- or 23-nucleotide spacer sequence in the opposite direction, followed by a conserved nonamer sequence. The Steen team established that cleavage only occurs between RSSs of different spacer lengths, preventing, for instance, V_H-to-V_H joining, by the same spacer lengths.

The recombinase complex binds two RSSs to create double-strand breaks (DSBs) in the DNA structure during the G1 phase of the cell cycle, bringing the two sites together via a synapse (Bredemeyer et al., 2008). The complex causes a nick in the 5' (phosphorylated) end of the RSS. The other strand of DNA is cleaved in the corresponding spot, leaving a 3' hydroxyl group exposed at the other strand. At this point, a gap is formed between a *coding end* and the *signal end*. A phosphodiester bond joins the *coding end* strands to create an intermediate *hairpin* structure (Van Gent et al., 1996). The cleavage process occurs at each of the two RSSs. Interestingly, the RAG proteins alone, even in a non-cellular system, were found sufficient for carrying out these steps by McBlane et al. (1995). The mechanism is curious in that these breaks usually occur in a malignant context.

Non-homologous end joining (NHEJ) triggers repair mechanisms, otherwise neoplasm or death of the cell would ensue. The two coding ends, one within the D_H cluster and one within the J_H cluster, need to be connected. In NHEJ, a DNA-dependent protein kinase–catalytic subunit ($DNA\text{-}PK_{cs}$), the Artemis specific nuclease, DNA ligase IV and other components are recruited.

Firstly, the DSB is detected by the Ku heterodimer (the ∼70kDa and ∼80kDa Ku70 and Ku80 polypeptides). Ku, identified first as an autoimmune antigen, contains a

ring structure near the C-terminus which binds to the major groove at the site of the DSB and a conserved sequence at the extreme C-terminus. This sequence recruits the DNA-PKcs (Radhakrishnan and Lees-Miller, 2017), a large (4000 amino-acid) serine/threonine protein kinase from the phosphatidylinositol-3-kinase family which comprises many signalling proteins involed, for example, in cell survival (Hartley et al., 1995). The DNA-PK_{cs} dimerises by creating a synapse with another DNA-PK_{cs} at the other coding end. Only after this synapse is formed, auto- and transphosphorylation occurs between the two units of the polypeptide (Hammel et al., 2010). Publications refer to the Ku heterodimer and the catalytic subunit collectively as DNA-PK.

Subsequently, the Artemis endonuclease is recruited via its C-terminal region by the DNA-PK_{cs} dimer. Artemis is a useful piece of cellular machinery: reported to be 'the only vertebrate nuclease capable of opening DNA hairpins' (Chang and Lieber, 2016), the N-terminal nuclease in Artemis has been observed to tackle a number of structures, including Y-shaped DNA structures, loops, bubbles and others, achieving to open them when in association with DNA-PK_{cs} (Ma et al., 2005). Ijspeert et al. (2011) describe reductions both in B- and T-cell counts in two patients with different null mutations in the gene for Artemis. In addition to supressed V(D)J recombination, the absence of Artemis increased susceptibility to ionising radiation, which was employed to introduce double-strand breaks in DNA.

A motif in the C-terminal region of Artemis binds the DNA-binding domain of DNA ligase IV (LigIV; an enzyme with the ability to connect DNA strands via a phosphodiester bond), which occurs in complex with the DNA repair protein XRCC4. (Calsou et al. /2003/ confirm, via immunoprecipitation studies, the sequentiality of the steps in NHEJ, showing that recruitment of the XRCC4-LigIV complex is dependent upon the assembly of the Ku-DNA-PK_{cs} complex.) XRCC4 functions as a scaffolding protein, with the ability to form multimeric structures with other units of XRCC4 or other proteins involved in the NHEJ process (reviewed by Williams et al., 2014). XRCC4 forms a super-helical structure which organises into filaments, where the XRCC4-like factor (XLF) binds into hydrophobic pockets of XRCC4 via a key leucine residue (Hammel et al., 2011). LigIV contains a tandem of two breast cancer susceptibility gene C-terminal (BRCT) domains, also occurent in other proteins involved in DNA repairs, which binds to XRCC4, as found by Critchlow et al. (1997) via common immunoprecipitation of the two. At this stage, LigIV is able to rejoin the two coding ends and create a recombined DJ_H segment. Figure 7 gives an overview of NHEJ.

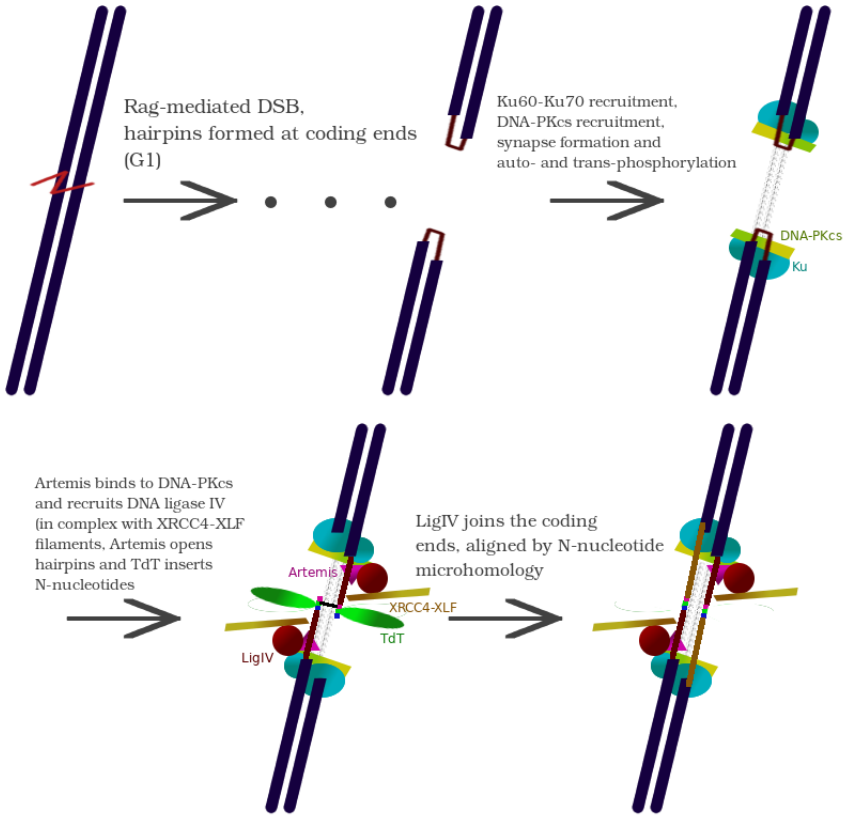

Figure 7. *A schematic representation of double-strand breaks and non-homologous end joining in the context of V(D)J recombination (self, 2018).*

The NHEJ model is a versatile approach to repairing double-strand breaks in the cell. An additional feature exclusive to lymphocytes, however, comes into play to increase the repertoire of possible outcomes which arise from the recombination. The terminal deoxynucleotidyl transferase (TdT) is a curious type of DNA polymerase, with the ability to elongate a 3' overhang, with no need for a second strand as a template (Chang, 1971). TdT was found to attach 'extra' nucleotides in a stochastic manner to both coding and signal ends of DNA (Lieber et al., 1988). The enzyme shows a consistent preference for dGTP and dCTP, which can be regulated by a variety of metal cations (Chang and Bollum, 2009). Let us focus again on the coding ends trimmed by Artemis-DNA-PK_{cs} as intermediates in the NHEJ process. When the hairpin structure at a coding end is trimmed, the hairpin does not open symmetrically at the tip (where it was sealed) but rather at one strand closer to the D- or J-segment. Therefore, upon opening, the hairpin structure unravels and reveals a sequence which is, by principle, palindromic, termed *P-nucleotides* (reviewed by Motea and Berdis, 2010). If this occurs in the context of V(D)J recombination, TdTs attach deoxynucleoside triphosphates to either Artemis-trimmed coding end, extending a non-templated (N) region at the 3' overhang at either side. Mickelsen et al. (1999) state the average number of added deoxynucleotide triphosphates lies in the range from 2 to 5. After the creation of N regions, the two regions on either of the coding ends

align homologically over a short segment which happens to fit with respect to the classical base-pairing rules. Overhangs which interfere with the homology are removed by the Artemis–DNA-PK_{cs} complex or other exonucleases involved (Motea and Berdis, 2010).

Adjacent to the V, D and J gene segments are C (constant) segments, responsible for the immunoglobulin *isotype*. Different isotypes have distinct roles in antigen-specific response. The repertoire of immunoglobulin isotypes expands tremendously during class-switch recombination later on (Horejsi et al., 2017).

3.4 From pro-B cell to pre-B cell

The transition to a pre-B cell, which expresses a pre-B cell receptor (IgH with a light chain surrogate), is marked, among other processes, by increasing IgH expression and an eventual rise in surface CD19 expression. Pre-B cells may then be gated based on a $CD34^{-}CD19^{+}CD10^{+}CD38^{+}TdT^{-}$ phenotype (Bagwell et al., 2015; Bendall et al., 2014).

After successful somatic D-J regrouping, the process is repeated to bring the V_H segment to DJ_H in a late pro-B cell (Alt et al., 1984). With a VDJ_H segment ready for transcription, the pro-B cell can proceed by expressing an immunoglobulin heavy chain (IgH), which is an integral part of a B-cell receptor. Kitamura and Rajewsky (1992) investigated the mechanism of *allelic exclusion*, whereby a successful IgH rearrangement process at one chromosome halts V-DJ recombination at the other chromosome. In addition to a lack of signalling via a functional pre-BCR, they identified mRNA coding for a viable membrane form of the heavy chain of isotype M as a trigger for the exclusion. Lutz et al. (2011), using flow cytometry for quantifications of cells from mutant mice, observed realiable response upon accumulation of μH mRNA that was not targeted by nonsense-mediated mRNA decay (NMD), which eliminates mRNA with premature termination codons. As demonstrated in the paper, stable yet noncoding mRNA can be created to avoid NMD, yet still suppress V-DJ recombination at the other chromosome, effectively halting development of the cell.

Allelic exclusion ensures a single clonotype of antibodies produced per B-cell. Barreto and Cumano (2000) found, using FACS with multiple anti-IgM and other stains, that only about 1 in 10 000 murine splenic B-cells showed dual IgH expression. This was presumably due to the inability of the first mRNA to form a functional pre-BCR receptor or due to synchronous expression at both chromosomes. The fascinating mechanism of exclusion based on mRNA seems to explain the relative effectiveness of the process.

Similar to a mature B-cell receptor, the pre-BCR needs two heavy chain units, two light chain units and other co-receptor proteins to assemble. The μHC (heavy chain) is joined by a surrogate ψLC (light chain) and an Igα/Igβ (CD79a/CD79b) transmembrane heterodimer. The surrogate light chain is a non-covalently bound complex of the V_{preB} and λ_5 proteins, with λ_5 binding to a μH C-subunit via a cystein residue (reviewed by

LeBien, 2000). Takeshi and Reth (1990) showed transfecting deficient cells to express these two proteins will enable IgH expression, where the IgH would otherwise be retained in the endoplasmic reticulum.

Upon assembly, the complete pre-BCR causes signalling via multiple protein-tyrosine kinases (more in Benschop, 1999). A B-cell specific linker molecule is of considerable interest as a modulator of signalling. BLNK (or SPL-65), a target of the Pax5 transcription factor, is responsible for pre-BCR signalling and has been identified as the proximal factor in developmental arrest due to Pax5 deficiencies (Schebesta et al., 2002).

3.5 Toward the immature B-cell

Transition toward an immature B-cell is accompanied by a $CD19^+$, $CD20^+$ phenotype and down-regulation of the CD10 and CD38 markers. Immature B-cells express surgace IgM which include physiologic light chains (Bagwell et al., 2015). Recombination of Ig light chain gene segments, contingent upon pre-BCR signalling, occurs in a similar fashion to IgH recombination. However, D-segments are absent (a V-J recombination occurs) and the mechanism of terminating recombination differs. Two types of Ig light chains are encoded: Igκ and Igλ. First, the V- and H-segments at the *IGK* locus undergo rearrangements until such mRNA is transcribed that is able to associate with the μH chain, creating a complete surface immunoglobulin. However, Igκ production fails at both chromosomes, Igλ production commences. This was first postulated upon observing leukaemic cells as a model by Alt el al. (1980) and corroborated by Asenbauer et al. (1999) who noted that the transcriptional enhancer for Igλ has limited accessibility to a DNase, as opposed to the Igκ enhancer. Hence, in addition to allelic exclusion, an *isotypic exclusion* (κ versus γ) takes place, resulting in a majority of BCRs only containing a single light-chain isotype. However, Giachino et al. (1995) used FACS to confirm an existent κ^+/λ^+ phenotype and indicated that allelic inclusion may also result in dual expression of the same isotype. Xu (2006) reports on dual light-chain expression in some B-cell chronic lymphocytic leukaemias, observed via flow cytometry.

With a complete B-cell receptor complex, comprised of surface IgM (sIgM) and Igα/Igβ, the immature B-cell is tested for autoreactivity prior to leaving the marrow.

3.6 Testing for autoreactivity

Multiple mechanisms are involved in establishing central B-cell tolerance. Immature IgM^+ B-lymphocytes are subjected to processes which weed out self-reactive cells. Wardemann et al. (2003) list some features of the complementarity-determining region 3 (CDR3; determinant of antibody specificity) in surface Ig which serve as indication of self-reactivity. These include increased length of CDR3 as well as increased share of positively-charged

amino acids. The portion of cells with such a phenotype decreased with progression toward mature B-cells; furthermore, the team used ELISA to test reactivity with various cell lysates and human cell antigens. Their conclusion was to estimate that between 55 and 75% of immature B-cells express self-reactive surface Ig.

3.6.1 Receptor editing

Gay et al. (1993) published a seminal paper on receptor editing: examining the fate of autoreactive immature B-cells in transgenic mice, the team found the cells to show delayed expression of *newly* rearranged IgL chains. Cytometry confirmed that the newly edited BCRs would not bind self-antibodies, when the original mutant BCRs did. Further results pointed toward the cells' ability to re-express activator genes for the reassembly of the V(D)J recombination machinery upon the multivalent cross-linking of sIgM when encountering a self-antigen (Tiegs et al., 1993). The possibility of multiple IgL rearrangements at this stage is documented (Yamagami et al., 1999). In addition, most of the primary rearrangements result in the use of the 1^{st} and 2^{nd} J$\varkappa$ regions, with multiple J and V regions left intact for potential secondary rearrangements. Autoreactive mutants were observed to retain the same number of B-cells as wild-type subjects (Pelanda et al., 1997).

As the Tiegs team reminds us, immature B-cells have the potential to express as many as 2 H chains and 6 L chains simultaneously, only barred by allelic and isotypic exclusion, lending some credence to an editing model, as opposed to mere deletion of self-reactive cells. This line of reasoning proves sensible: Chen et al. (1995) identified an analogous secondary rearrangement mechanism in IgH in transgenic autoreactive cells. However, these rearrangements are rather more complex than in the case of IgL: two rearrangements (D-to-J and V-to-DJ) were originally needed for a functional IgH segment recombination (detailed above). Fanning et al. (1998) describe a heptamer sequence adjacent to the 3' end of most V_H segments, which is used often in the secondary rearrangement as a new RSS, enabling replacement of the V_H segment by a different one. Moreover, a less-frequent *direct* V_H-to-J_H joining, leading to omission of the D_H segment, was observed (Koralov et al., 2006). However interesting this mechanism of secondary rearrangement of IgH is, Zhang et al. (2003) found that only about 5 % of functional human IgH sequences bore a footprint of this process.

Crucially, Schram et al. (2008) postulate a 'threshold' for antigen stimulation (via *tonic signalling*) needed to halt *Rag* expression and receptor editing. The group worked with transgenic mice HEL-specific Ig and GFP-tagged RAG2 (allowing for quantification via flow cytometry). They showed that immunodeficient subjects with decreased classical BCR signalling (which occurs via cross-linking and recruitment of the Btk kinase) showed no decrease in receptor editing, refuting a link between acute BCR cross-linking and editing in the given conditions. Furthermore, when stimulated by the self-antigen *in vitro*,

both control and immunodeficient cells showed minute RAG2–GFP expression. Surface activation markers CD69 and CD86 were monitored, as Ag was introduced in 'pulses': RAG2 expression showed a notable delay when compared to these markers; in addition, its eventual up-regulation was similar in the control and immunodeficient cell populations. In contrast to this, cells deficient in the *lyn* Src family kinase (responsible for Igα and Igβ phosphorylation) saw elevated levels of RAG2-GFP, suggesting a distinct mechanism in accordance with a basal signalling model.

3.6.2 Clonal deletion

Once thought to be the prevalent mechanism for establishing central B-cell tolerance, clonal deletion of autoreactive cells was inferred from studies using early transgenic mouse models. For instance, Nemazee and Buerki (1989) observed the apparent loss of transgenic self-specific B-cells, originally targeting MHC class I (a histo-compatibility molecule), both in the marrow and spleen. The team attributed this effect to prevalent clonal deletion of those B-cells in the marrow which reacted to the autologous antigen.

A subsequent study by Hartley et al. (1993) described an initial stage of developmental arrest followed by hypothesised apoptosis. However, shift in forward and side scatter (in cytometry) or chromatin condensation, indicative of apoptosis, was, conspicuously, not detected in the immature B-cell population. Due to the inability of this team to capture cell death, they stated at the time it '[could not] be excluded that the B cells may persist but lose all cell surface markers' or that other cells encompass the apoptotic vesicles. While it is apparent that some clonal deletion takes place, it appears to mostly occur in cases of failure to generate a non-autoreactive secondary rearrangement of the BCR (described above and reviewed by Pelanda and Torres, 2012).

3.6.3 Anergy

The result of immature B-cell receptor testing is not a dichotomy. Some antigen binding with avidity brings about a state of anergy. A key study by Goodnow et al. (1988) described transgenic immature B-cells targeted at hen-egg lysozyme (HEL), which was also expressed. In defiance of a supposed clonal deletion mechanism for eliminating self-reactive cells, a number of them appeared in secondary lymphoid organs, with increased internalisation of IgM. Benschop et al. (2001) studied a transgenic immature B-cell population binding a bone marrow autoantigen; whereas no receptor editing was observed, activation markers (including the B220 activation marker) were seen in flow cytometry. However, the nascent mature B-cells showed no signalling upon BCR stimulation. Later, Merrell et al. (2006) identified anergic B-cells in the periphery in wild-type mouse and Quach et al. (2011) complemented these findings, investigating a population of human naïve B-cells with abnormally low IgM expression and impaired signalling.

3.7 Revealing a developmental trajectory

To get a comprehensive picture of B-cell development, let us evaluate some studies which take advantage of single-cell technology to reveal the trajectory which the aforementioned progenitors follow.

An example of an early study of transitioning populations is a work by Terstappen et al. (1991). The group conducted a five-dimensional analysis of bone marrow cells. In an attempt to order transitioning populations of B-lymphoid cells, they identified a steady increase in CD38 and slowed decrease in CD34, with a significant CD10 expression increase characteristic for B-lymphopoiesis. These transitioning populations could be visualised using combinations of two-dimensional dot-plots, thanks to the relatively small number of parameters measured.

As to more recent studiues, mass cytometry is of great use. In a 2014 project, Bendall et al. developed a 44-parameter marker panel and an algorithm for both clustering populations within ∼200,000-cell sets and ordering the measured events along a continuum (shown in figure 8), dubbed *Wanderlust*. A graph-based method, Wanderlust finds neighbours of cells along a developmental trajectory not by simply identifying the closest different data point, but by constructing a sensible path from point A to point B which covers all the intermediate points on its way. Wanderlust creates multitudes of possible trajectories and constructs a consensus path, which mostly avoids this 'short-circuiting' (false connections). The trajectory *must* be non-branching, rendering B-cell lymphopoiesis a good candidate for testing the algorithm.

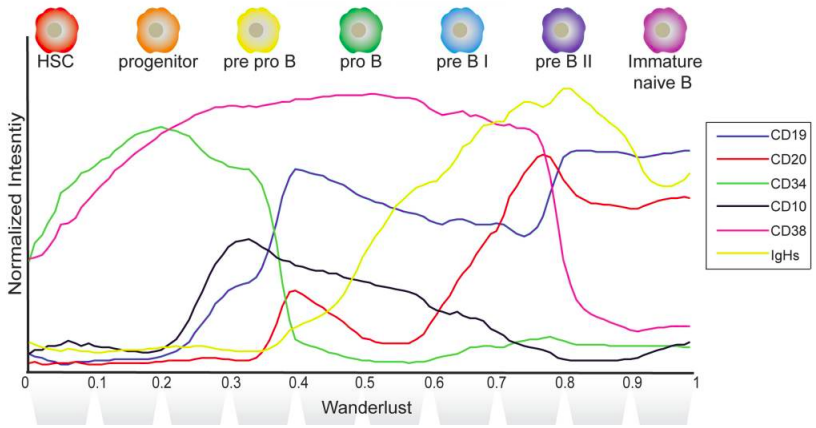

Figure 8. *Intensity of signal from some CD markers was normalised and used to characterise B-cell development. Adapted from Bendall et al. (2014).*

Interestingly, the non-parametric* Spearman correlation coefficient (ρ) never dropped under $\pm$ 0.97 (with maximum at 1) with the exclusion of any one parameter from the anal-

*Because the statistical test is non-parametric, the rank (i.e. ordering) of datapoints along a pseudotime continuum is important, rather than an exact measure of their position on this scale.

ysis. The choice of a correct starting point for the analysis, however, is important, due to the zero branching assumption. The paths generated by Wanderlust were consistent with quantitative PCR (qPCR) analysis focused on changes in IgH as a metric of development. The results were also consistent across different individuals.

After constructing the developmental trajectory, the group investigated the significance of IL-7 signalling, marked by activation of STAT5 kinase (used as an intracellular marker) in humans. Wanderlust found a single phenotypically distinct population of cells (about 0.007% of total) which saw an approximately five-fold increase in STAT5 activation, showing a narrow timeframe where IL-7 signalling functions as a checkpoint for IgH rearrangement. Importantly, STAT5 activation was not a known parameter to Wanderlust, showing an extraordinary ability to reveal checkpoints in cell development from juxtaposing abundant data from many markers that are examined simultaneously.

Bagwell et al. (2015) investigated the sequentiality of some stages in B-lymphopoiesis. They point to the problematic nature of synthesising knowledge from cord blood and bone marrow samples, quantifying cell populations based on manual gates or studying malignancies to draw conclusions regarding healthy lymphopoiesis. The team used the *GemStone* software for *probability state modelling* (PSM): a complex algorithm which calculates the most probable ordering of events.

The team first ran an 'unconstrained' analysis, which identified 6 'control definition points', which seem critical in the progression. This way, a sharp down-regulation of CD34, subsequent slight increase in CD45, a second CD45 up-regulation coinciding with sharp CD20 up-regulation and the virtually simultaneous drop in CD38 and CD10 were identified; developmental stages were then mapped onto a progression axis (see figure 9). Interestingly, the *B3* stage (corresponding to an immature B-cell stage) was either subtle or not observed in some samples upon individual inspection.

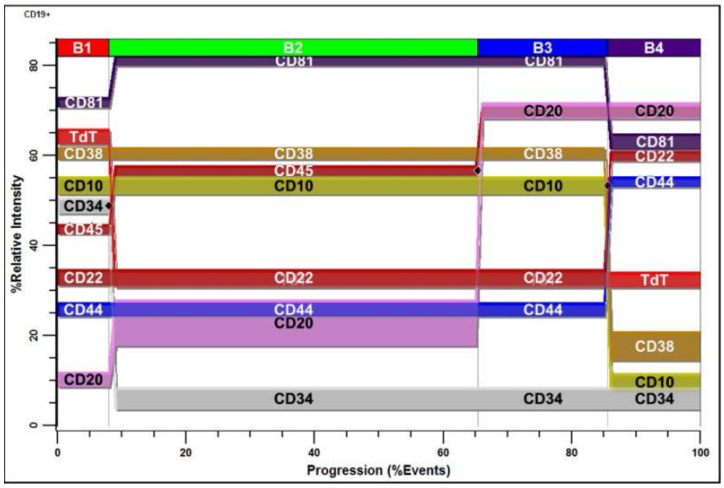

Figure 9. *An average model of lymphopoiesis after B-lineage commitment, using relative intensities of various markers. Statistically inferred developmental stages (B1 to B4) are*

shown (Bagwell et al., 2015).

The group also presents the ability of GemStone to constrain the analysis, allowing the user to force some crucial changes in marker expression to coincide on the progression axis. This provides an exciting framework to build robust models of lymphopoiesis and use them to compare aberrant data from patients and identify the aetiology of their illness.

4 Conclusion

B-cell lymphopoiesis is a convoluted and critical process in the immune system. Using various markers and stains, high-throughput single-cell technology can help identify distinct populations of cells and serve as a powerful tool in explaining the processes within B-cell development.

From supression of various oncogenes, proper adhesion signalling, transcription of genes for commitment to a particular myeloid or lymphoid lineage to somatic rearrangement processes for assembling a functional antigen receptor, haematopoiesis is made up of nuanced trajectories. Many molecular mechanisms within these trajectories are not understood in their entirety.

With the advent of mass cytometry coupled with novel smart approaches to automatised clustering of populations, new opportunities to aggregate data regarding haematopoiesis are emerging. If our knowledge regarding healthy blood development is robust, objective models can be built to serve as templates for diagnosis of malignancies and a better understanding of disease aetiology can be established. New opportunities for targeted therapy can be explored. Also, differences between mouse models and human patients can be investigated in depth.

Cluster analyses and algorithms which reveal complicated developmental trajectories, combined with good understanding of genetics, might bring about an era in which artificial intelligence provides sound judgement in diagnostics and treatment of leukaemias and immunodeficits, steered by medical professionals, biologists and informaticians who oversee the process.

5 List of abbreviations

7-AAD: 7-aminoactinomycin D

Ab: antibody

BCR: B-cell receptor

BLNK: B-cell linker

BM: bone marrow

BRCT: breast cancer susceptibility gene C-terminal

BrdU: bromodeoxyuridine

Btk: Bruton's tyrosine kinase

CB: cord blood

CD: cluster of differentiation

CDR3: complementarity-determining region 3

CLP: common lymphoid progenitor

CSC: cancer stem cell

CXCL12: C-X-C motif chemokine 12; stromal cell-derived factor 1

CXCR4: C-X-C chemokine receptor type 4

CyTOF: Time of Flight Mass Cytometry

γc: γ chain [of IL-7R]

DAPI: 4',6-diamidino-2-phenylindole

dCTP: deoxycytidine triphosphate

dGTP: deoxyguanosine triphosphate

DNA: deoxyribonucleic acisd

DNA-PKcs: DNA-dependent protein kinase – catalytic subunit

DSB: double-strand breaks

EBF1: early B-cell factor-1

EBF1: early B-cell factor-1

ELISA: enzyme-linked immunosorbent assay

FACS: fluorescence-activated cell sorting

FITC: fluorescein isothiocyanate

Flt-3: ms like tyrosine kinase 3

FSC: forward scatter

GFP: green fluorescent protein

GM-CSF: granulocyte-macrophage colony-stimulating factor

HEL: hen-egg lysozyme

HSA: heat-stable antigen

HSC: haematopoietic stem cell

ICP: inductively coupled plasma

Ig: immunoglobulin

IgH: immunoglobulin heavy chain

IgL: immunoglobulin light chain

IL-7: interleukin-7

IL-7R: interleukin-7 receptor

IL-7Rα: alpha chain of the interleukin-7 receptor

LigIV: DNA ligase IV

LMPP: lymphoid-primed multipotent progenitor

Ly6D: lymphocyte antigen 6 family D

m/z: mass-to-charge ratio

MHC: major histocompatibility complex

MPP: multipotent progenitor

mRNA: messenger RNA

NHEJ: non-homologous end-joining

NMD: nonsense-mediated mRNA decay

PCA: principal component analysis

PCR: polymerase chain reaction

PI: propidium iodide

pRB: retinoblastoma protein

PSM: probability state modelling

RAG: recombination-activating gene

RNA: ribonucleic acid

RSS: recombination signal sequence

SOM: self-organising map

SSC: side scatter

STAT: signal transducer and activator of transcription

T_c: cytotoxic T-lymphocytes

TdT: terminal deoxynucleotidyl transferase

T_h: helper T-lymphocytes

TSLP: thymic stromal lymphopoietin

t-SNE: *t*-distributed stochastic neighbour embedding

VCAM-1: vascular cell adhesion molecule 1

XLF: XRCC4-like factor; Cernunnos

XRCC4: X-ray repair cross-complementing protein 4

6 Bibliography

Secondary sources are marked with an asterisk (*).

* Adan, A., Alizada, G., Kiraz, Y., Baran, Y., and Nalbant, A. (2017). *Flow cytometry: basic principles and applications.* Critical Reviews in Biotechnology, 37(2):163–176.

Adolfsson, J., Månsson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C. T., Bryder, D., Yang, L., Borge, O. J., Thoren, L. A., Anderson, K., Sitnicka, E., Sasaki, Y., Sigvardsson, M., and Jacobsen, S. E. W. (2005). *Identification of Flt3+lympho-myeloid stem cells lacking erythro-megakaryocytic potential: A revised road map for adult blood lineage commitment.* Cell, 121(2):295–306.

Alt, F. W., Enea, V., Bothwell, a. L., and Baltimore, D. (1980). *Activity of multiple light chain genes in murine myeloma cells producing a single, functional light chain.* Cell, 21(1):1–12.

Alt, F. W., Yancopoulos, G. D., Blackwell, T. K., and Tonegawal, S. (1984). *Ordered rearrangement of immunoglobulin heavy chain variable region segments.* 3(6):1209–1219.

Amir, E. A. D., Davis, K. L., Tadmor, M. D., Simonds, E. F., Levine, J. H., Bendall, S. C., Shenfeld, D. K., Krishnaswamy, S., Nolan, G. P., and Pe'Er, D. (2013). *ViSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia.* Nature Biotechnology, 31(6):545–552.

Asenbauer, H., Combriato, G., and Klobeck, H. G. (1999). *The immunoglobulin lambda light chain enhancer consists of three modules which synergize in activation of transcription.* European Journal of Immunology, 29(2):713–724.

Bagwell, C. B. and Adams, E. G. (1993). *Fluorescence Spectral Overlap Compensation for Any Number of Flow Cytometry Parameters.* Annals of the New York Academy of Sciences.

Bagwell, C. B., Hill, B. L., Wood, B. L., Wallace, P. K., Alrazzak, M., Kelliher, A. S., and Preffer, F. I. (2015). *Human B-cell and progenitor stages as determined by probability state modeling of multidimensional cytometry data.* Cytometry Part B - Clinical Cytometry, 88(4):214–226.

Bain, G., Maandag, E. C., Izon, D. J., Amsen, D., Kruisbeek, a. M., Weintraub, B. C., Krop, I., Schlissel, M. S., Feeney, a. J., and van Roon, M. (1994). *E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements.* Cell, 79:885–892.

Bandura, D. R., Baranov, V. I., Ornatsky, O. I., Antonov, A., Kinach, R., Lou, X., Pavlov, S., Vorobiev, S., Dick, J. E., and Tanner, S. D. (2009). *Mass Cytometry: A Novel Technique for Real-Time Single Cell Multi-Target Immunoassay Based on Inductively Coupled Plasma Time of Flight Mass Spectrometry.* Analytical Chemistry, 81(16):6813–6822.

Barreto, V. and Cumano, A. (2000). *Frequency and Characterization of Phenotypic Ig Heavy Chain Allelically Included IgM-Expressing B Cells in Mice.* The Journal of Immunology, 164(2):893–899.

Becker, J. S., Matusch, A., Depboylu, C., Dobrowolska, J., and Zoriy, M. V. (2007). *Quantitative Imaging of Selenium, Copper, and Zinc in Thin Sections of Biological Tissues (Slugs - Genus Arion) Measured by Laser Ablation Inductively Coupled Plasma Mass Spectrometry.* Analytical Chemistry, 79(16):6074–6080.

Bendall, S. C., Davis, K. L., and Amir, E.-a. D. (2014). *Single-Cell Trajectory Detection Uncovers Progression and Regulatory Coordination in Human B cell Development.* Cell, 27(4):339–351.

Benschop, R. J., Aviszus, K., Zhang, X., Manser, T., Cambier, J. C., Wysocki, L. J., and Medical, N. J. (2001). *Activation and Anergy in Bone Marrow B Cells of a Novel Immunoglobulin Transgenic Mouse that Is Both Hapten Specific and Autoreactive.* Immunity, 14:33–43.

* Benschop, R. J. and Cambier, J. C. (1999). *B cell development: Signal transduction by antigen receptors and their surrogates.* Current Opinion in Immunology, 11(2):143–151.

Bigos, M., Baumgarth, N., Jager, G. C., Herman, O. C., Nozaki, T., Stovel, R. T., Parks, D. R., and Herzenberg, L. A. (1999). *Nine Color Eleven Parameter Immunophenotyping Using Three Laser Flow Cytometry.* Cytometry, 45:36–45.

* Bratthauer, G. L. (2010). *The Avidin-Biotin Complex (ABC) Method and Other Avidin-Biotin Binding Methods.* Methods in Molecular Biology, 588:257–270.

Bredemeyer, A. L., Helmink, B. A., Innes, C. L., Calderon, B., Lisa, M., Mahowald, G. K., Gapud, E. J., Walker, L. M., Collins, J. B., Weaver, B. K., Mandik-nayak, L., Schreiber, R. D., Allen, P. M., Michael, J., Paules, R. S., Bassing, C. H., Sleckman, B. P., Group, C., Sciences, E. H., and Carolina, N. (2008). *DNA double strand breaks activate a multi-functional genetic program in developing lymphocytes.* Nature, 456(7223):819–823.

Burk, A. S., Monzel, C., Yoshikawa, H. Y., Wuchter, P., Saffrich, R., Eckstein, V., Tanaka, M., and Ho, A. D. (2015). *Quantifying Adhesion Mechanisms and Dynamics of Human Hematopoietic Stem and Progenitor Cells.* Scientific Reports, 5:1–8.

Callebaut, I. and Mornon, J. P. (1998). *The V(D)J recombination activating protein RAG2 consists of a six-bladed propeller and a PHD fingerlike domain, as revealed by sequence analysis.* Cellular and Molecular Life Sciences, 54(8):880–891.

Calsou, P., Delteil, C., Frit, P., Drouet, J., and Salles, B. (2003). *Coordinated assembly of Ku and p460 subunits of the DNA-dependent protein kinase on DNA ends is necessary for XRCC4-ligase IV recruitment.* Journal of Molecular Biology, 326(1):93–103.

* Chang, H. H. and Lieber, M. R. (2016). *Structure-Specific nuclease activities of Artemis and the Artemis: DNA-PKcs complex.* Nucleic Acids Research, 44(11):4991–4997.

Chang, L. M. S. (1971). *Development of terminal deoxynucleotidyl transferase activity in embryonic calf thymus gland.* Biochemical and Biophysical Research Communications, 44(1):124–131.

Chang, L. M. S. and Bollum, F. J. (2009). *Multiple Roles of Divalent Cation in the Terminal Deoxynucleotidyltranferase Reaction.* The Journal of Biological Chemistry, 34(March):91–96.

Chappaz, S., Flueck, L., Farr, A. G., Rolink, A. G., and Finke, D. (2007). *Increased TSLP availability restores T- and B-cell compartments in adult IL-7-deficient mice.* Blood, 110(12):3862–3870.

Chen, C., Nagy, Z., Luning, E., and Weigert, M. (1995). *Immunoglobulin Heavy Chain Gene Replacement: A Mechanism of Receptor Editing.* Immunity, 3:747–755.

* Corfe, S. A. and Paige, C. J. (2012). *The many roles of IL-7 in B cell development; Mediator of survival, proliferation and differentiation.* Seminars in Immunology, 24(3):198–208.

Critchlow, S. E., Bowater, R. P., and Jackson, S. P. (1997). *Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV.* Current Biology, 7(8):588–598.

Curbelo, R., Block, J., Shapiro, M., and Searle, G. D. (1976). *A generalized machine for automated flow cytology system design.* The Journal of Histochemistry and Cytochemistry, pages 388–395.

DeKoter, R. P., Lee, H.-j., Singh, H., and Biology, C. (2002). *PU.1 Regulates Expression of the Interleukin-7 Receptor in Lymphoid Progenitors.* Immunity, 16:297–309.

Dias, S., Månsson, R., Gurbuxani, S., Sigvardsson, M., and Kee, L. (2008). *E2A proteins promote development of lymphoid-primed multipotent progenitors.* Immunity, 29(2):217–227.

* Edling, C. E. and Hallberg, B. (2007). *c-Kit–A hematopoietic cell essential receptor tyrosine kinase.* International Journal of Biochemistry and Cell Biology, 39(11):1995–1998.

Fulwyler, M. J. (1965). *Electronic Separation of Biological Cells by Volume.* Science, 150.

Funk, P. E., Kincade, P. W., and Witte, P. L. (1994). *Native associations of early hematopoietic stem cells and stromal cells isolated in bone marrow cell aggregates.* Blood, 83(2):361–9.

Garvy, B. A. and Riley, R. L. (1994). *IFN-gamma abrogates IL-7-dependent proliferation in pre-B cells, coinciding with onset of apoptosis.* Immunology, 81(3):381–8.

Gay, D., Saunders, T., Camper, S., and Alerts, E. (1993). *Receptor Editing: An Approach by Autoreactive B Cells to Escape Tolerance.* The Journal of Immunology, 177:999–1008.

* Glish, G. L. and Vachet, R. W. (2003). *The basics of mass spectrometry in the twenty-first century.* Nature Reviews Drug Discovery, 2(2):140–150.

Hammel, M., Rey, M., Yu, Y., Mani, R. S., Classen, S., Liu, M., Pique, M. E., Fang, S., Mahaney, B. L., Weinfeld, M., Schriemer, D. C., Lees-Miller, S. P., and Tainer, J. A. (2011). *XRCC4 protein interactions with XRCC4-like factor (XLF) create an extended grooved scaffold for DNA ligation and double strand break repair.* Journal of Biological Chemistry, 286(37):32638–32650.

Hammel, M., Yu, Y., Mahaney, B. L., Cai, B., Ye, R., Phipps, B. M., Rambo, R. P., Hura, G. L., Pelikan, M., So, S., Abolfath, R. M., Chen, D. J., Lees-Miller, S. P., and Tainer, J. A. (2010). *Ku and DNA-dependent protein kinase dynamic conformations and assembly regulate DNA binding and the initial non-homologous end joining complex.* Journal of Biological Chemistry, 285(2):1414–1423.

Hartley, K. O., Gell, D., Smith, G. C. M., Zhang, H., Divecha, N., Connelly, M. A., Admon, A., Lees-Miller, S. P., Anderson, C. W., and Jackson, S. P. (1995). *DNA-Dependent Protein Kinase Catalytic Subunit: A Relative of Phosphatidylinositol 3-Kinase and the Ataxia Telangiectasia Gene Product.* Cell, 82:849–856.

Hartley, S. B., Cooke, M. P., Fulcher, D. A., Harris, A. W., Cory, S., Basten, A., and Goodnow, C. C. (1993). *Elimination of Self-Reactive B Lymphocytes Proceeds in Two Stages: Arrested Development and Cell Death.* Cell, 72:325–335.

Heavey, B., Charalambous, C., Cobaleda, C., and Busslinger, M. (2003). *Myeloid lineage switch of Pax5 mutant but not wild-type B cell progenitors by C/EBP-alpha and GATA factors.* The EMBO Journal, 22(15):3887–3897.

Herzenberg, L. A., Sweet, R. G., and Herzenberg, L. A. (1976). *Fluorescence-activated Cell Sorting.* Scientific American, 273(2):78–83.

Hewitt, S. L., Yin, B., Ji, Y., Chaumeil, J., Marszalek, K., Tenthorey, J., Salvagiotto, G., Steinel, N., Ramsey, L. B., Farrar, M. A., Sleckman, B. P., Schatz, D. G., Bassing, C. H., Skok, J. A., Haven, N., Biocenter, V., Universitaire, C., and Haven, N. (2009). *RAG1 and ATM coordinate monoallelic recombination and nuclear positioning of immunoglobulin loci.* Nature Immunology, 10(6):655–664.

Hinton, G. and Roweis, S. (2002). *Stochastic Neighbor Embedding.*

Hogg, W. R. and Coulter, W. H. (1971). *United States Patent 3,557,352.*

Holmes, M. L., Carotta, S., Corcoran, L. M., and Nutt, S. L. (1992). *Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein.* The EMBO Journal, 11:327–333.

* Horejsi, V., Bartunkova, J., Brdcka, T., and Spisek, R. (2017). *Základy imunologie, 6. vydání.*

Ijspeert, H., Lankester, A. C., Van Den Berg, J. M., Wiegant, W., Van Zelm, M. C., Weemaes, C. M. R., Warris, A., Pan-Hammarström, Q., Pastink, A., Van Tol, M. J. D., Van Dongen, J. J. M., Van Gent, D. C., and Van Der Burg, M. (2011). *Artemis splice defects cause atypical SCID and can be restored in vitro by an antisense oligonucleotide.* Genes and Immunity, 12(6):434–444.

Inlay, M. A., Bhattacharya, D., Sahoo, D., Serwold, T., Seita, J., Karsunky, H., Plevritis, S. K., Dill, D. L., and Weissman, I. L. (2009). *Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development.* Genes & Development, 23(20):2376–2381.

Inomata, M., Takahashi, S., Harigae, H., Kameoka, J., Kaku, M., and Sasaki, T. (2006). *Inverse correlation between Flt3 and PU.1 expression in acute myeloblastic leukemias.* Leukemia Research, 30(6):659–664.

Jacobsen, K., Kravitz, J., Kincade, P. W., and Osmond, D. G. (1996). *Adhesion receptors on bone marrow stromal cells: in vivo expression of vascular cell adhesion molecule-1 by reticular cells and sinusoidal endothelium in normal and gamma-irradiated mice.* Blood, 87(1):73–82.

Johnson, S. E., Shah, N., Panoskaltsis-Mortari, A., and LeBien, T. W. (2005). *Murine and Human IL-7 Activate STAT5 and Induce Proliferation of Normal Human Pro-B Cells.* The Journal of Immunology, 175(11):7325–7331.

Kaneko, H., Ariyasu, T., Inoue, R., Fukao, T., Kasahara, K., Teramoto, T., Matsui, E., Hayakawa, S., and Kondo, N. (1998). *Expression of Pax5 gene in human haematopoietic cells and tissues: Comparison with immunodeficient donors.* Clinical and Experimental Immunology, 111(2):339–344.

Katz, F. E., Tindle, R., Sutherland, D. R., and Greaves, M. F. (1985). *Identification of a membrane glycoprotein associated with haemopoietic progenitor cells.* Leukemia Research, 9(2):191–198.

Kikushige, Y., Yoshimoto, G., Miyamoto, T., Iino, T., Mori, Y., Iwasaki, H., Niiro, H., Takenaka, K., Nagafuji, K., Harada, M., Ishikawa, F., and Akashi, K. (2008). *Human Flt3 Is Expressed at the Hematopoietic Stem Cell and the Granulocyte/Macrophage Progenitor Stages to Maintain Cell Survival.* The Journal of Immunology, 180(11):7358–7367.

Kitamura, D. and Rajewsky, K. (1992). *Targeted disruption of mu chain membrane exon causes loss of heavy-chain allelic exclusion.* Nature, 355:472–475.

Kohonen, T. (1998). *Exploration of Very Large Databases by Self-Organizing Maps.* Neurocomputing, 21(1-3):1–6.

Koralov, S. B., Novobrantseva, T. I., Ehlich, A., Ko, J., and Rajewsky, K. (2006). *Antibody Repertoires Generated by VH to JH Joining.* Immunity, (July):43–53.

Kosak, S. T., Skok, J. A., Medina, K. L., Riblet, R., Le Beau, M. M., Fisher, A. G., and Singh, H. (2002). *Subnuclear Compartmentalization of Immunoglobulin Loci During Lymphocyte Development.* Science, 296(April):158–163.

Kozmik, Z., Wang, S., Dorfler, P., Adams, B., and Busslinger, M. (1992). *The Promoter of the CD19 Gene Is a Target for the B-Cell-Specific Transcription Factor BSAP.* Molecular and Cellular Biology, 12(6):2662–2672.

* LeBien, T. W. and Tedder, T. F. (2008). *B lymphocytes: how they develop and function.* Blood, 112(5):1570–1580.

Lechner, J., Chen, M., Hogg, R. E., Toth, L., Silvestri, G., Chakravarthy, U., and Xu, H. (2015). *Alterations in Circulating Immune Cells in Neovascular Age-Related Macular Degeneration.* Scientific Reports, 5:1–10.

Lemay, P. and Collyn-d'Hooghe, M. (1984). *Flow Cytophotometric and Time-lapse Cinematographic Study of Human Cells Infected by Adenovirns Type 2 Wild-type and Two DNA-negative Temperature-sensitive Mutants.* Journal of General Virology, (1984):1419–1423.

Levin, S. D., Koelling, R. M., Friend, S. L., Isaksen, D. E., Ziegler, S. F., Perlmutter, R. M., and Farr, A. G. (1999). *Thymic stromal lymphopoietin: a cytokine that promotes the development of IgM+ B cells in vitro and signals via a novel mechanism.* The Journal of Immunology, 162(2):677–83.

Lieber, M. R., Hesse, J. E., Mizuuchi, K., and Gellert, M. (1988). *Lymphoid V(D)J recombination: nucleotide insertion at signal joints as well as coding joints.* Proceedings of the National Academy of Sciences of the United States of America, 85(22):8588–92.

Lugli, E., Pinti, M., Nasi, M., Troiano, L., Ferraresi, R., Mussi, C., Salvioli, G., Patsekin, V., Robinson, J. P., Durante, C., Cocchi, M., and Cossarizza, A. (2007). *Subject Classification Obtained by Cluster Analysis and Principal Component Analysis Applied to Flow Cytometric Data.* Cytometry Part A, 69(A):659–676.

Ma, Y., Schwarz, K., and Lieber, M. R. (2005). *The Artemis:DNA-PKcs endonuclease cleaves DNA loops, flaps, and gaps.* DNA Repair, 4(7):845–851.

Matthews, A. G., Kuo, A. J., Ramón-Maiques, S., Han, S., Champagne, K. S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D. N., Walter, K. L., Utz, P. J., Shi, Y., Kutateladze, T. G., Yang, W., Gozani, O., and Oettinger, M. A. (2007). *RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination.* Nature, 450(7172):1106–1110.

McBlane, J. F., Van Gent, D. C., Ramsden, D. A., and Cuomo, A. (1995). *Cleavage at a V(D)J Recombination Signal Requires Only RAG1 and RAG2 Proteins and Occurs in Two Steps.* Cell, 83(D):387–395.

Merrell, K. T., Benschop, R. J., Gauld, S. B., Aviszus, K., Decote-ricardo, D., Wysocki, L. J., Cambier (2006). *Identification of Anergic B Cells within a Wild-Type Repertoire.* Immunity, (December):953–962.

Mickelsen, S., Snyder, C., Trujillo, K., Bogue, M., Roth, D. B., and Meek, K. (1999). *Modulation of terminal deoxynucleotidyltransferase activity by the DNA-dependent protein kinase.* Journal of Immunology, 163(2):834–843.

Möhle, R., Bautz, F., Rafii, S., Moore, M. A. S., Brugger, W., Kanz, L., and Sdf, T. (2013). *The Chemokine Receptor CXCR-4 Is Expressed on CD34+ Hematopoietic Progenitors and Leukemic Cells and Mediates Transendothelial Migration Induced by Stromal Cell-Derived Factor-1.* Blood, pages 4523–4530.

Molnár, Á. and Georgopoulos, K. (1994). *The Ikaros Gene Encodes a Family of Functionally Diverse Zinc Finger DNA-Binding Proteins.* Molecular and Cellular Biology, 14(12):8292–8303.

Morrison, S. J. and Weissman, I. L. (1994). *The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype.* Immunity, 1(8):661–673.

* Motea, E. A. and Berdis, A. J. (2010). *Terminal deoxynucleotidyl transferase: The story of a misguided DNA polymerase.* Biochimica et Biophysica Acta, 1804(5):1151–1166.

Murphy, R. F. and Chused, T. M. (1984). *A Proposal for a Flow Cytometric Data File Standard.* Cytometry, 555:553–555.

Nemazee, D. A. and Buerki, K. (1989). *Clonal deletion of B lymphocytes in a transgenic mouse bearing anti-MHC class I antibody genes.* Nature, 337(February).

Nomizu, T., Hayashi, H., Hoshino, N., Tanaka, T., Kitagawa, K., and Kaneco, S. (2002). *Determination of zinc in individual airborne particles by inductively coupled plasma mass spectrometry with digital signal processing.* Journal of Analytical Atomic Spectrometry, 17:592–595.

Nomizu, T., Kaneco, S., Tanaka, T., Ito, D., Kawaguchi, H., and Vallee, B. T. (1994). *Determination of Calcium Content in Individual Biological Cells by Inductively Coupled Plasma Atomic Emission Spectrometry.* Analytical Chemistry, 66(19):3000–3004.

Oettinger, M. A., Schatz, D. G., Gorka, C., and Baltimore, D. (1990). *RAG-1 and RAG-2, Adjacent Genes That Synergistically Activate V(D)J Recombination.* Science.

Okuno, Y., Iwasaki, H., Huettner, C. S., Radomska, H. S., Gonzalez, D. a., Tenen, D. G., and Akashi, K. (2002). *Differential regulation of the human and murine CD34 genes in hematopoietic stem cells.* Proceedings of the National Academy of Sciences of the United States of America, 99(9):6246–51.

* Ormerod, M. G. (2008). *Flow Cytometry – A Basic Introduction.* URL: *http://flowbook.denovosoftware.com/.*

Osawa, M., Hanada, K., Hamada, H., and Nakauchi, H. (1996). *Long-Term Lymphohematopoietic Reconstitution by a Single CB34-Low/Negative Hematopoietic Stem Cell.* Science, 4947(1 995).

Parks, R., Loken, R., and Herzenberg, A. (1977). *Two-color immunofluorescence.* The Journal of Histochemistry and Cytochemistry, pages 899–907.

Parrish, Y. K., Baez, I., Milford, T., Benitez, A., Galloway, N., Rogerio, J. W., Sahakian, E., Kaagoda, M., Huang, G., Hao, Q., Sevilla, Y., Barsky, L. W., Zielinska, E., Price, M. A., Wall, N. R., Dovat, S., and Payne, K. J. (2009). *IL-7 Dependence in Human B Lymphopoiesis Increases During Progression of Ontogeny from Cord Blood to Bone Marrow.* The Journal of Immunology, 49(18):1841–1850.

Pelanda, R., Schwers, S., Sonoda, E., Torres, R. M., Nemazee, D., and Rajewsky, K. (1997). *Receptor Editing in a Transgenic Mouse Model: Site, Efficiency, and Role in B Cell Tolerance and Antibody Diversification.* Immunity, 7:765–775.

* Pelanda, R. and Torres, R. M. (2012). *Central B-Cell Tolerance: Where Selection Begins.* Cold Springs Harbor Perspectives in Biology, pages 1–15.

Peschon, J. J., Morrisey, P. J., Grabstein, K. H., Ramsdell, F. J., Maraskovsky, E., Gliniak, B. C., Park, L. S., Ziegler, S. F., Williams, D. E., Ware, C. B., Meyer, J. D., and Davison, B. L. (1994). *Early Lymphocyte Expansion Is Severely Impaired in Interleukin 7 Receptor-deficient Mice.* The Journal of Experimental Medicine, 180(November):6–11.

Pongubala, J. M., Northrup, D. L., Lancki, D. W., Medina, K. L., Treiber, T., Bertolino, E., Thomas, M., Grosschedl, R., Allman, D., and Singh, H. (2008). *Transcription factor EBF restricts alternative lineage options and promotes B cell fate commitment independently of Pax5.* Nature Immunology, 9(2):203–215.

* Pribyl, J. A. R. and LeBien, T. W. (1996). *Interleukin 7 independent development of human B cells.* Proceedings of the National Academy of Sciences of the United States of America, 93(19):10348–53.

Puel, A., Ziegler, S. F., Buckley, R. H., and Leonard, W. J. (1998). *Defective IL7R expression in T-B+NK+ severe combined immunodeficiency.* Nature Genetics, 20 (December).

Quach, T. D., N., M.-O., Adlowitz, D. G., Silver, L., H., Y., C., W., Milner, E. C. B., and Sanz, I. (2011). *Anergic Responses Characterize a Large Fraction of Human Autoreactive Naïve B Cells Expressing Low Levels of Surface IgM.* The Journal of Immunology, 186(8):4640–4648.

Radhakrishnan, S. K. and Lees-Miller, S. P. (2017). *DNA requirements for interaction of the C-terminal region of Ku80 with the DNA-dependent protein kinase catalytic subunit (DNA-PKcs).* DNA Repair.

Reynaud, D., Demarco, I., Reddy, K., Schjerven, H., Bertolino, E., Chen, Z., Smale, S. T., Winandy, S., and Singh, H. (2009). *Regulation of B cell fate commitment and immunoglobulin VH gene rearrangements by Ikaros.* Nature Immunology, 9(8):927–936.

Roederer, M. (2001). *Spectral compensation for flow cytometry: Visualization artifacts, limitations, and caveats.* Cytometry, 45(3):194–205.

Sayegh, C., Jhunjhunwala, S., Riblet, R., and Murre, C. (2005). *Visualization of looping involving the immunoglobulin heavy-chain locus in developing B cells.* Genes & Development, pages 322–327.

Schatz, D. G., Oettinger, M. A., and Baltimore, D. (1989). *The V(D)J Recombination Activating Gene, RAG-1.* Cell, 59(D):1035–1048.

Schebesta, M., Pfeffer, P. L., and Busslinger, M. (2002). *Control of pre-BCR signaling by Pax5-dependent activation of the BLNK gene.* Immunity, 17(4):473–485.

Scheeren, F. A., Van Lent, A. U., Nagasawa, M., Weijer, K., Spits, H., Legrand, N., and Blom, B. (2010). *Thymic stromal lymphopoietin induces early human B-cell proliferation and differentiation.* European Journal of Immunology, 40(4):955–965.

Schram, B. R., Tze, L. E., Ramsey, L. B., Liu, J., Najera, L., Vegoe, A. L., Hardy, R. R., Keli, L., Farrar, M. A., Behrens, T. W., and Alerts, E. (2008). *B Cell Receptor Basal Signaling Regulates Antigen-Induced Ig Light Chain Rearrangements.* The Journal of Immunology.

Schweitzer, B. L. and DeKoter, R. P. (2004). *Analysis of Gene Expression and Ig Transcription in PU.1/Spi-B-Deficient Progenitor B Cell Lines.* The Journal of Immunology, 172(1):144–154.

Seamer, L. C., Bagwell, C. B., Barden, L., Redelman, D., Salzman, G. C., and Murphy, R. F. (1997). *Proposed New Data File Standard for Flow Cytometry, Version FCS 3.0.* Cytometry, 122:118–122.

Simmons, P. J. and Torok-Storb, B. (1991). *CD34 expression by stromal precursors in normal human adult bone marrow.* Blood, 78(11):2848–2853.

Sonn, C., Cho, J., Kim, J., Kang, M., Lee, J., and Kim, J. (2016). *Detection of circulating tumor cells in patients with non-small cell lung cancer using a size-based platform.* Oncology Letters, pages 2717–2722.

Steen, S. B., Gomelsky, L., and Roth, D. B. (1996). *The 12/23 rule is enforced at the cleavage step of V(D)J recombination.* Genes to Cells, (D):543–553.

Steininger, A., Möbs, M., Ullmann, R., Köchert, K., Kreher, S., Lamprecht, B., Anagnostopoulos, I., Hummel, M., Richter, J., Beyer, M., Janz, M., Klemke, C.-D., Stein, H., Dörken, B., Sterry, W., Schrock, E., Mathas, S., and Assaf, C. (2011). *Genomic loss of the putative tumor suppressor gene E2A in human lymphoma.* The Journal of Experimental Medicine, 208(8):1585–1593.

Sun, L., Liu, A., and Georgopoulos, K. (1996). *Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development.* The EMBO Journal, 15(19):5358–69.

Takeshi, T. (1990). *The Products of Pre-B Cell-specific Genes (lambda5 and VpreB).* The Journal of Experimental Medicine, 172(September):973–976.

Terstappen, L. W., Huang, S., Safford, M., Lansdorp, P. M., and Loken, M. R. (1991). *Sequential generations of hematopoietic colonies derived from single nonlineage-committed CD34+CD38- progenitor cells.* Blood, 77(6):1218–27.

Tiegs, S. L., Russell, D. M., and Nemazee, D. (1993). *Receptor Editing in Self-reactive Bone Marrow B Cells.* The Journal of Experimental Medicine, pages 1009–1020.

Urbánek, P., Wang, Z. Q., Fetka, I., Wagner, E. F., and Busslinger, M. (1994). *Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5 BSAP.* Cell, 79(5):901–912.

Van Dilla, M. A., Trujillo, T. T., Mullaney, P. F., and Coulter, J. R. (1969). *Cell Microfluorometry: A Method for Rapid Fluorescent Measurement.* Science, (March).

Van Gassen, S., Callebaut, B., Van Helden, M. J., Lambrecht, B. N., Demeester, P., Dhaene, T., and Saeys, Y. (2015). *FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data.* Cytometry Part A, 87(7):636–645.

Van Gent, D. C., Mizuuchi, K., and Gellert, M. (1996). *Similarities Between Initiation of V(D)J Recombination and Retroviral Integratio*n. Science, 271(March):1–3.

Van Unen, V., Höllt, T., Pezzotti, N., Li, N., Reinders, M. J., Eisemann, E., Koning, F., Vilanova, A., and Lelieveldt, B. P. (2017). *Visual analysis of mass cytometry data by hierarchical stochastic neighbour embedding reveals rare cell types.* Nature Communications, 8(1).

Voßhenrich, C. A., Cumano, A., Müller, W., Di Santo, J. P., and Vieira, P. (2003). *Thymic stromal-derived lymphopoietin distinguishes fetal from adult B cell development.* Nature Immunology, 4(8):773–779.

Wardemann, H. (2003). *Predominant Autoantibody Production by Early Human B Cell Precursors.* Science, 301(5638):1374–1377.

* Weber, L. M. and Robinson, M. D. (2016). *Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data.* Cytometry Part A, 89(12):1084–1096.

* Wilkerson, M. J. (2012). *Principles and Applications of Flow Cytometry and Cell Sorting in Companion Animal Medicine.* Veterinary Clinics of North America – Small Animal Practice, 42(1):53–71.

* Williams, G. J., Hammel, M., Radhakrishnan, S. K., Ramsden, D., Lees-Miller, S. P., and Tainer, J. A. (2014). *Structural insights into NHEJ: building up an integrated picture of the dynamic DSB repair super complex, one component and interaction at a time.* DNA Repair, 17:110–120.

Williamson, D. H. and Fennell, D. J. (1974). *Apparent dispersive replication of yeast mitochondrial DNA as revealed by density labelling experiments.* Molecular & General Genetics, 131(3):193–207.

Wold, S., Esbensen, K., and Geladi, P. (1987). *Principal component analysis.* Chemometrics and Intelligent Laboratory Systems, 2(1-3):37–52.

Xu, D. (2006). *Dual surface immunoglobulin light-chain expression in B-cell lymphoproliferative disorders.* Archives of Pathology and Laboratory Medicine, 130(6):853–856.

Yamagami, Y., ten Boekel, E., Andersson, J., Rolink, A., and Melchers, F. (1999). *Frequencies of Multiple IgL Chain Gene Rearrangements in Single Normal or kappa-L Chain–Deficient B Lineage Cells.* Immunity, 11:317–327.

Zhang, Z., Zemlin, M., Wang, Y.-h., Munfus, D., Huye, L. E., Findley, H. W., Bridges, S. L., Roth, D. B., Burrows, P. D., and Cooper, M. D. (2003). *Contribution of VH Gene Replacement to the Primary B Cell Repertoire.* Immunity, 19:21–31.